Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile

Ministero della Transizione Ecologica

Ricerca di Sistema elettrico

Smart Energy in sistemi pubblici: analisi di affidabilità e qualificazione dei dati per ridurre le incertezze di sistema

Fabio Leccese, Mariagrazia Leccisi

ROMA TRE UNIVERSITÀ DEGLI STUDI

Report RdS/PTR2020/021

SMART ENERGY IN SISTEMI PUBBLICI: ANALISI DI AFFIDABILITÀ E QUALIFICAZIONE DEI DATI PER RIDURRE LE INCERTEZZE DI SISTEMA

Fabio Leccese, Mariagrazia Leccisi

Aprile 2021

Report Ricerca di Sistema Elettrico

Accordo di Programma Ministero dello Sviluppo Economico - ENEA
Piano Triennale di Realizzazione 2019-2021 - II annualità
Obiettivo: Tecnologie
Progetto: Tecnologie per la penetrazione efficiente del vettore elettrico negli usi finali
Work package: Local Energy District
Linea di attività: 30 - Ampliamento del PELL in relazione agli edifici pubblici scolastici e analisi affidabilistica di Smart Road
Responsabile del Progetto: Claudia Meloni, ENEA
Responsabile del Work package: Claudia Meloni, ENEA

Il presente documento descrive le attività di ricerca svolte all'interno dell'Accordo di collaborazione "*Smart Energy in Sistemi Pubblici: analisi di affidabilità e qualificazione dei dati per ridurre le incertezze di sistema"*
Responsabile scientifico ENEA: Ing. Fabio Moretti
Responsabile scientifico Roma Tre – Dipartimento di Scienze Dott. Ing. Ph.D. RTI Fabio Leccese

Sommario

Nell'ambito del progetto "Smart Energy in Sistemi Pubblici: analisi di affidabilità e qualificazione dei dati per ridurre le incertezze di sistema", l'attività svolta dall'unità dell'Università degli Studi "Roma Tre" ha riguardato quattro filoni principali:

- Il miglioramento della piattaforma PELL dal punto di vista dell'esperienza utente, compresa la gestione di anomalie e dei problemi risultanti dall'utilizzo del sistema, in continuità con quanto effettuato nell'anno precedente ed il supporto all'implementazione di Specifici Key Performance Index.
- Supporto alla progettazione di un portale per il censimento degli edifici pubblici scolastici, dal punto di vista energetico analogamente a quanto già esistente per l'illuminazione pubblica
- Un'attività informatica di supporto alla gestione di notifiche di funzionalità dei server attivi.
- Lo studio di algoritmi utili al monitoraggio dei consumi di energia elettrica

1 Introduzione

Nell'ambito del piano triennale della ricerca 2019-2021 per il sistema elettrico nazionale, per il quale l'ENEA ha predisposto il piano triennale di realizzazione, PTR 2019-2021, il Dipartimento di Scienze dell'Università degli Studi "Roma Tre" è stato interessato per una attività di ricerca dal titolo "Smart Energy in Sistemi Pubblici: Analisi di Affidabilità e Qualificazione dei Dati per Ridurre le Incertezze di Sistema".
All'interno di questo quadro generale, l'Università, durante il secondo anno (2020), è stata coinvolta nelle seguenti attività che ricadono all'interno del working package 1 LA 30:

- PELL IP,
- PELL edifici,
- Attività informatica,
- Ricerca di algoritmi di efficienza energetica.

Mentre la ~~quart~~a attività è stato un task di esclusiva competenza dell'Università, le altre due ci hanno visto coinvolti in specifiche attività a supporto di quella principale guidata da ENEA.
In particolare:

- PELL- IP: ci siamo occupati di test e verifica di nuove funzionalità e di quelle già in essere sia per la fase statica sia per la fase dinamica.

-PELL EDIFICI: siamo stati di supporto nella fase di analisi e progettazione di un nuovo portale per il censimento degli edifici pubblici.

- Attività informatica: gestione di notifiche di funzionalità dei server attivi.

In relazione alla quarta attività, è importante anticipare che l'attività di ricerca di algoritmi per l'efficienza energetica e il monitoraggio degli impianti non è fine a se stessa, ma è stata immaginata e sviluppata per soddisfare le specifiche esigenze provenienti dal PELL, ed in particolare da quella sezione del lavoro che ricade sotto il nome di KPI.

Riassumendo quindi, le prime tre attività si sono configurate come attività di integrazione e supporto a quanto sviluppato da altri in seno al progetto, mentre la terza, di nostra esclusiva competenza, si è giovata dello sviluppo della prima attività per poterne studiare un miglioramento sulla base dei suoi dati.

2 PELL-IP

Il PELL è un progetto complesso che si pone l'obiettivo di riorganizzare la visione della città alla luce di un modello di management efficiente ed efficace [1,2].

Il PELL IP si occupa di illuminazione pubblica che è considerato un tema chiave per lo sviluppo di una città moderna. Attualmente, questo settore è gestito prevalentemente dalle amministrazioni pubbliche che, spesso per mancanza di sensibilità, lo rilegano in secondo piano rispetto ad altri temi di maggiore impatto sulla attività di gestione. Questo spesso produce inefficienza del sistema di illuminazione, dovuto in prevalenza all'età degli impianti che non vengono aggiornati con le nuove tecnologie disponibili sul mercato e, talvolta, alla loro scarsa manutenzione che tipicamente implica costi spesso poco sostenibili da parte delle amministrazioni. D'altro canto, questa situazione, apre ampi margini di miglioramento in termini di prestazioni e conseguenti risparmi di risorse energetiche ed economiche.

Una illuminazione pubblica definibile "smart", utilizza tecnologie rispettose dell'ambiente (es. led) e delle risorse del pianeta, diminuendo al contempo l'impatto energetico, i costi di manutenzione e migliorando l'efficienza e la funzionalità della stessa illuminazione.

Un altro punto importante, oltre alla manutenzione fisica dell'impianto al fine di apportarne miglioramenti, risiede nella valutazione delle prestazioni funzionali ed energetiche che permettono di capire dove intervenire al fine di rispondere alle domande "dove serve, quanto serve e per quanto tempo serve" in modo da massimizzare l'efficienza dell'impianto.

Questo favorisce una importante riqualificazione tecnologica, economica, sociale ed urbanistica della città.

A tali scopi, il primo passo previsto dal sistema è il censimento degli impianti di illuminazione pubblica, attraverso caricamento di XML condivisi con i vari fornitori di energia, di specifici tools per valutare le prestazioni illuminotecniche oppure per effettuare valutazioni economico/finanziarie ed eventuali investimenti per la riqualificazione.

Il progetto è orientato a varie tipologie di utenti, come Amministrazioni locali e centrali, Gestori, Operatori settoriali e anche cittadini.

In seno al progetto, noi siamo stati di supporto alle attività ENEA, attraverso attività di test e controllo di anomalie riscontrate durante l'utilizzo del sistema e, ove possibile, con interventi diretti volti al miglioramento dell'applicazione o allo sviluppo di funzionalità [3].

Di seguito alcuni esempi di modalità di test e modifiche da noi effettuati in merito a nuove funzionalità che sono state oggetto di test approfondito prima di essere rese disponibili agli utenti finali, in quanto implementate ex-novo.

L'elenco non è esaustivo di tutti i test che sono stati effettuati e di tutte le funzionalità, esistenti o nuove, che abbiamo verificato, però permette di capire il tipo di lavoro che abbiamo effettuato definibile da lessico tecnico come "valutazione della qualità del prodotto legata alla esperienza utente (user experience)" e lascia trasparire la mole di lavoro svolta.

2.1 Test accesso utenti

Ogni utente deve essere registrato sulla piattaforma, ed è classificato come:

- Comune: può inserire i dati dei propri impianti sulla piattaforma, usufruendo così dei servizi di diagnostica, monitoraggio dei consumi e indicatori di performance, al fine di valutare lo stato dell'impianto.
- Gestore: ha accesso al PELL al fine di caricare i dati relativi ai comuni che gestisce.
- Sviluppatore: può accedere alla piattaforma per avere una conoscenza puntuale e standardizzata del livello tecnologico, qualitativo e prestazionale degli impianti di illuminazione pubblica considerati. La

sua registrazione è vincolata ad una procedura definita “PELL Verified” che consiste in un badge virtuale che PELL rilascia ai soggetti che lo richiedono.

Gli utenti che desiderino inviare dati al PELL, devono completare una procedura online che verifica la loro capacità di produrre documenti “PELL compliant” da gestire sulla piattaforma. (Figura 1).

- Cittadino: un qualsiasi cittadino italiano può accedere al PELL e ai suoi moduli pubblici.

Figura 1-Dettaglio della pagina Web relativa alla Procedura per il conseguimento del badge PELL-Verified

I test di sistema servono a controllare che una specifica funzionalità sia stata implementata correttamente; in particolare ci siamo occupati di verificare il corretto funzionamento delle funzionalità di:

- Registrazione comune
- Registrazione gestore
- Registrazione sviluppatore
- Registrazione cittadino

In seguito all’inserimento dei dati nel form (Figura 2- Pagina Web del PELL IP contenete il form per la registrazione cittadino), il tasto “invia richiesta”, invia i dati all’applicazione che, in base al tipo di registrazione, inserisce il nuovo utente all’interno del sistema, salvandolo sul database.
L’utente viene definito di classe A (gestori, sviluppatori, comuni, Figura 3) o di classe B (generalmente i cittadini, Figura 4).

In seguito alla registrazione, il sistema deve mandare una email di conferma all’utente creato.
I test sono stati rivolti a tutte le fasi della funzionalità, in particolare:

- Corretta compilazione dei form per tutte le categorie di utenti con verifica di eventuali campi obbligatori/non obbligatori, con formati definiti (es. email) come indicato da specifiche.
- Verifica di assenza di errori javascript in fase di compilazione.
- Verifica di corretto salvataggio sul database nelle varie tabelle (registrazione_cittadino, registrazione_comune, registrazione_gestore, registrazione_sviluppatore).
- Verifica di assenza di utenti duplicati nel sistema.
- Verifica di invio email all’utente e verifica di correttezza delle informazioni presenti nelle email.

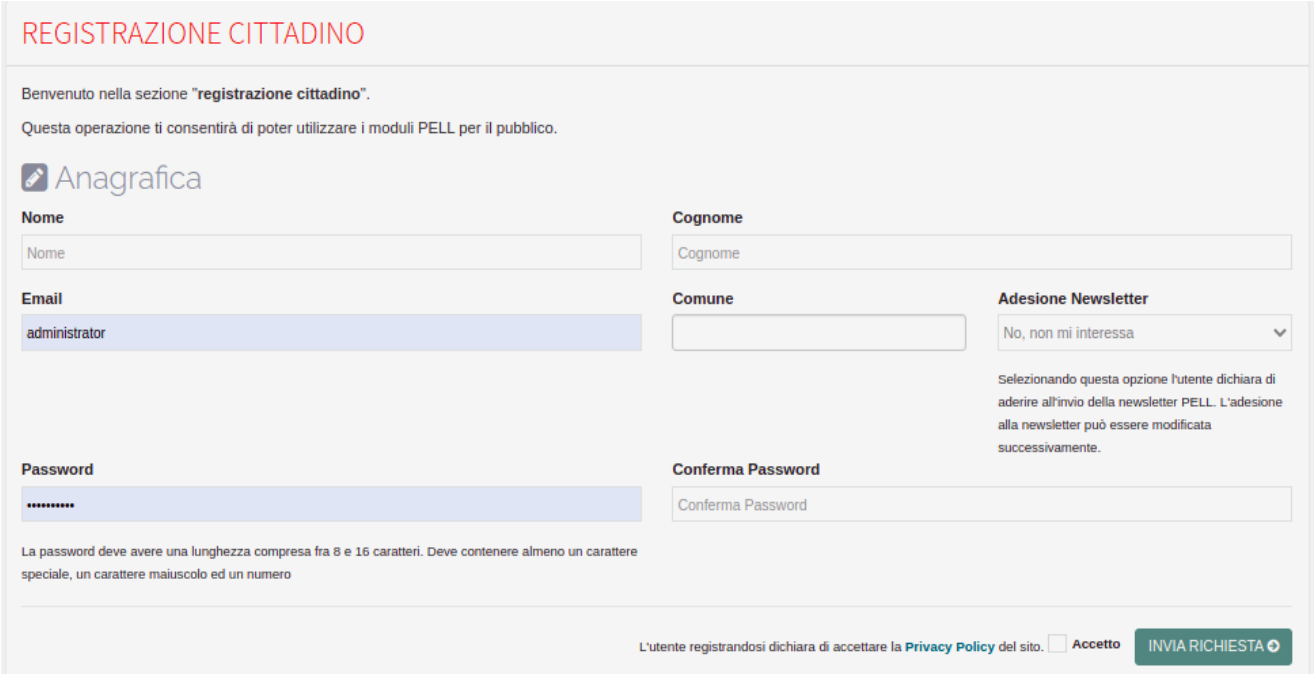

Figura 2- Pagina Web del PELL IP contenete il form per la registrazione cittadino. Pagine analoghe contenenti un form di inserimento dati sono presenti anche per la registrazione di comuni, gestori e sviluppatori

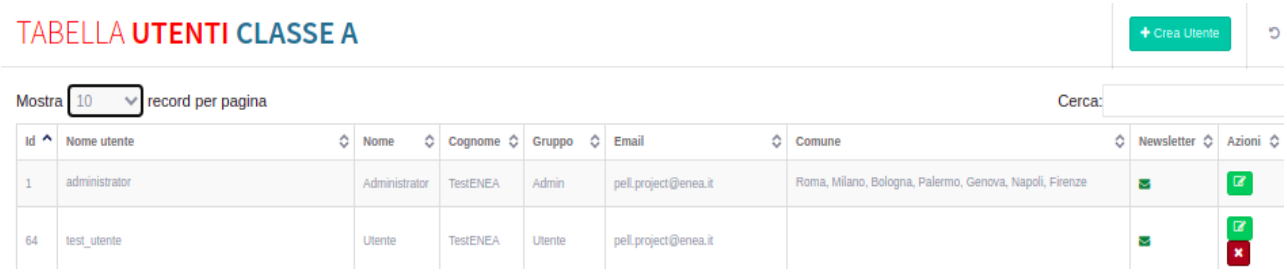

Figura 3- Tabella utenti di classe A, cioè quegli utenti che appartengono ai vari gruppi definiti sul pell, come Gestore, sindaco, sviluppatore ecc..

Figura 4 – tabella utenti classe B, cioè i cittadini che si registrano nel portale

2.2 *Test gestione e invio newsletter*

Sempre in merito alla gestione degli utenti, in seguito alla sua implementazione, è stata testata la funzionalità di invio delle newsletter agli utenti registrati.
In fase di registrazione, e comunque in qualsiasi momento, una volta effettuato l'accesso, un utente può selezionare/deselezionare un flag specifico per richiedere l'invio della newsletter all'email con cui si è registrato. La Figura 5 mostra il dettaglio della sezione del PELL IP "Info Utente" in cui Sono indicate la tipologia dell'utente (administrator, gestore ecc..) e il Flag "Invio Newsletter" che indica la possibilità per

l'utente di ricevere per email le newsletter. I test hanno verificato la selezionabilità/desezionabilità di questo flag da parte dell'operatore stesso.
Lo stesso flag è presente anche nel profilo dell'utente. La Figura 6 mostra il -Dettaglio della sezione del PELL IP relativa al "Profilo Utente", in cui è possibile verificare l'adesione all'invio delle newsletter. Anche da questa pagina è necessario poter selezionare o deselezionare l'adesione alla ricezione della newsletter.

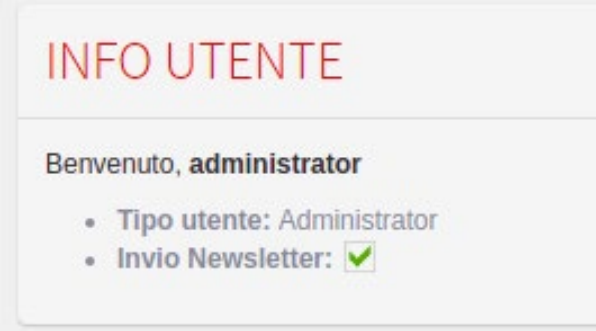

Figura 5-Dettaglio della sezione "Info Utente", relativa alle informazioni legate all'utente che ha effettuato il login nell'applicazione PELL IP.

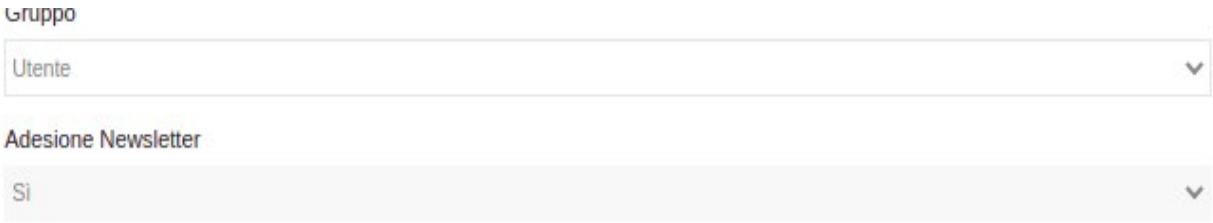

Figura 6-Dettaglio della sezione del PELL IP relativa al "Profilo Utente"

Essendo la funzionalità nuova, è stato necessario verificare ogni singolo aspetto, in particolare:

- Lato utente:
  - ✓ Verifica selezione/deselezione del flag in fase di iscrizione (per tutti i form di registrazione).
  - ✓ Verifica selezione/deselezione del flag durante l'accesso al portale con le proprie credenziali.
  - ✓ Verifica selezione/deselezione del flag attraverso modifica del profilo utente.

- Lato gestione delle newsletter:
  - ✓ Verifica della possibilità, da parte di un utente abilitato, di gestire le newsletter dal menu "Amministrazione". La Figura 7 mostra il link alla pagina di gestione delle newsletter, da cui è possibile creare, modificare ed eliminare una newsletter, nonché aggiungere o eliminare i destinatari della stessa.

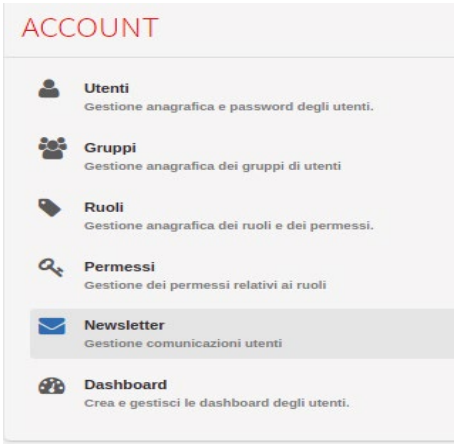

Figura 7 – Menu amministrazione in cui è stato inserito un link alla pagina di gestione delle newsletter

- ✓ Verifica della creazione di una newsletter attraverso la compilazione del form relativo (La Figura 8 mostra il form di creazione di una newsletter in cui è possibile allegare files e inserire destinatari, gruppi oppure indirizzi email specifici), accessibile attraverso il tasto "Crea Newsletter" nella pagina di gestione delle stesse, che comprende:
  - Corretta compilazione dei campi e acquisizione sulla tabella dedicata del database.
  - Corretta selezione e acquisizione di un allegato.
  - Corretta gestione della lista di email, relative ad altri utenti, a cui inviare le newsletter.

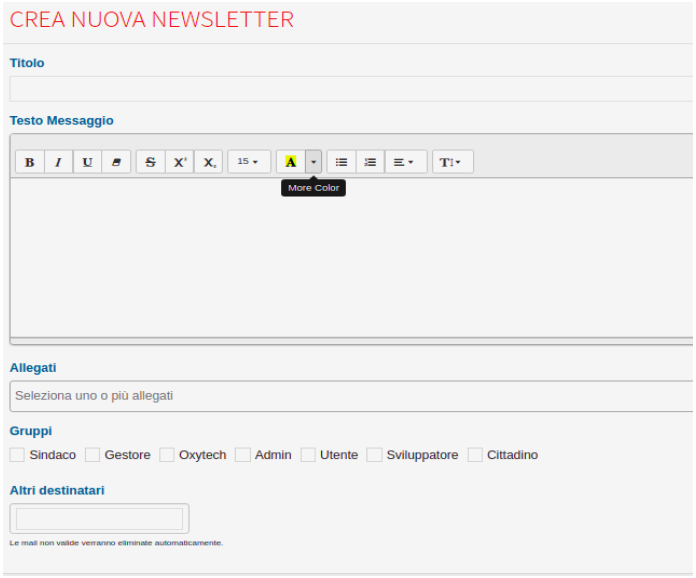

Figura 8-Form di creazione di una newsletter.

- ✓ Verifica della newsletter nella "Tabella Newsletter" e verifica degli allegati nella "Tabella Allegati" (Figura 9). Per ogni newsletter sono visibili alcune informazioni salienti come i destinatari e il titolo, e una lista di azioni che possono essere effettuate direttamente da questa pagina, come la modifica, l'eliminazione o l'invio).
- ✓ Verifica di inserimento di nuovi allegati nei formati ammessi e sua eventuale eliminazione.
- ✓ Verifica funzionalità dei tasti, relativi ad ogni newsletter, di "Modifica" di "Eliminazione" e di "Invio".
- ✓ Verifica di effettivo invio dell'email e della correttezza della newsletter ricevuta.

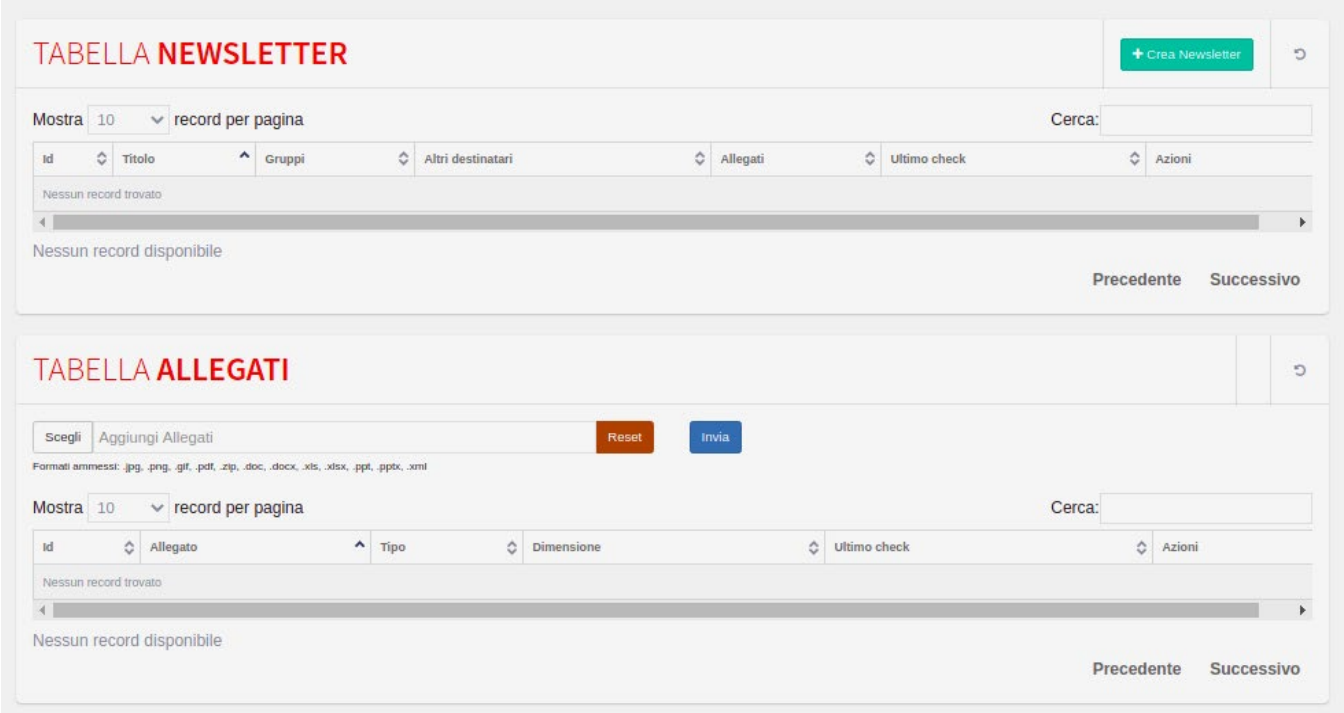

Figura 9- La lista delle newsletter create e degli allegati da inviare, presenti nella sezione "Newsletter", accessibile dal menu "Amministrazione".

2.3 *Test su gestione Cookies e privacy*

Sono stati effettuati dei test sulla funzionalità, obbligatoria per legge, di gestione dei cookies da parte dei visitatori del portale. Infatti è obbligatorio che un sito web dia la possibilità ad un visitatore di visualizzare i cookies utilizzati e di poter selezionare o deselezionare l'accettazione di determinati cookies. La Figura 10 mostra la richiesta di approvazione dei cookies. Per semplificare la gestione degli stessi, è stata prevista la possibilità di selezionare dei gruppi di cookies(necessari, preferenze, statistici, marketing). I cookies appartenenti al gruppo "Necessari" sono automaticamente selezionati, mentre altri cookies sono facoltativi e possono essere deselezionati secondo le necessità dell'utente.

La Figura 11-mostra il pannello di dettaglio dei cookies utilizzati dal sito, in cui vengono visualizzati i dettagli relativi ai cookies, in modo da favorire l'utente nella scelta consapevole di quelli da accettare e da rispondere alle esigenze di trasparenza indicate nella normativa di riferimento.

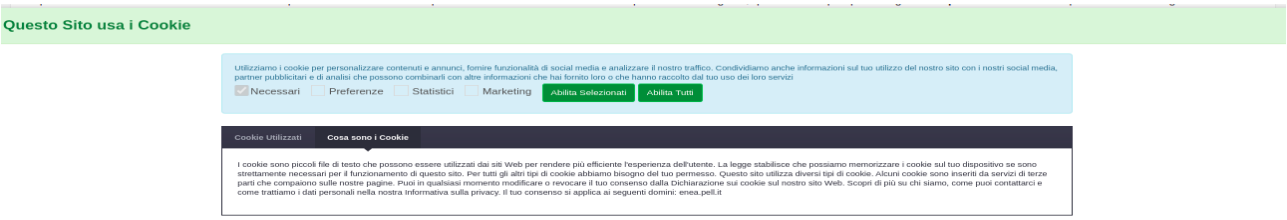

Figura 10-. Quando non già effettuato, all'accesso alla home del sito viene richiesta l'approvazione dei cookies

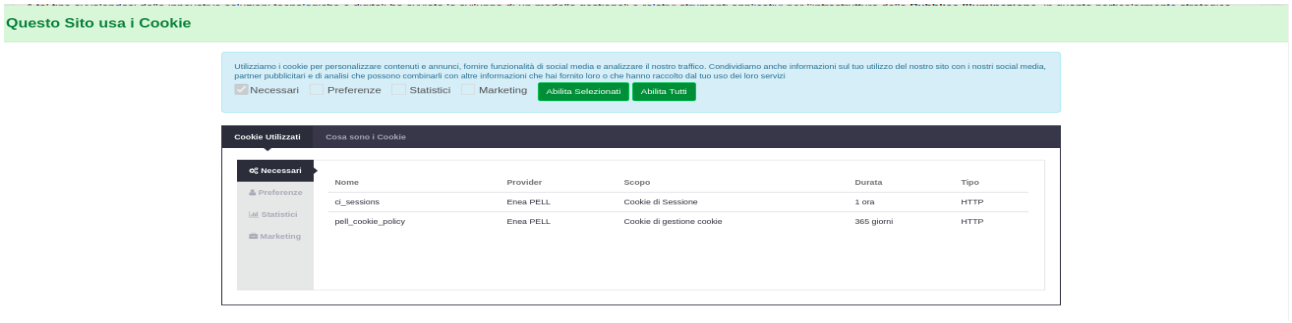

Figura 11- Pannello di dettaglio dei cookies utilizzati dal sito.

In tale contesto abbiamo riscontrato vari problemi di gestione di cookies, prevalentemente dovuti alle differenze di gestione degli stessi da parte di vari sistemi operativi.

I test sono stati volti dunque:

- A verificare la funzionalità dell'abilitazione/disabilitazione sui seguenti browser: Chrome, Firefox ed Edge.
- A verificare che quanto indicato nei popup fosse coerente con quanto previsto dalla normativa.
- A verificare la corretta gestione manuale dei popup, per modificare l'abilitazione ai cookies.

2.4 *Inserimento Area News*

Ove possibile e necessario, sono stati fatti anche interventi pratici sull'applicazione, un esempio in Figura 12 è l'inserimento di una pagina statica relativa ad un' "area news" in cui inserire varie informazioni di utilità. Il suo aggiornamento prevede l'inserimento del titolo e il testo lato codice nell'html.

Figura 12-Pagina "Area News".

A tale scopo è stata utile la struttura MVC (Model View Controller) di CodeIgniter, un framework PHP molto leggero che permette di creare web applications e che è stato utilizzato per lo sviluppo dell'applicazione.

Esso consente la divisione dell'applicazione secondo il pattern Model-View-Controller, separando di fatto:

- L'accesso al database (Model).
- L'interfaccia utente (View).
- La gestione dei dati del database, ed eventuali operazioni su di essi prima di essere passati alla view (Controller).

È stata quindi creata la pagina statica news.PHP nella cartella specifica destinata alle views; in questa cartella è presente una sottocartella "statica" che contiene tutte le pagine che non appartengono alle sezioni private del PELL e che sono dunque accessibili a chiunque utilizzi il portale anche se utente non autenticato.
In seguito alla creazione della pagina, è stata dunque inserita la chiamata ad essa da parte del web, attraverso il controller Welcome.PHP, che si occupa del semplice rendering (cioè la creazione e visualizzazione di tutti i vari elementi html di una pagina) della pagina, dal momento che essa non prevede alcun accesso al database La Figura 13Figura 13 mostra la funzione news() nel controller, che carica il titolo della pagina ed effefttua il rendering.

```
public function news() {
        $this->data['title'] = 'News';
        $this->render('enea/static/news');
}
```

Figura 13-La funzione news() nel controller Welcome.php

Sono dunque stati configurati i percorsi attraverso cui accedere alla pagina, aggiornando il file routes.PHP nelle configurazioni di CodeIgniter.
La Figura 14 mostra la modalità di definizione dei vari percorsi, mappando ogni controller tramite un "alias" che andrà a sostituire il link completo (enea/welcome/…).Nel nostro caso il browser invoca il controller "enea/welcome/background" attraverso l'alias "background".
Infine è stato aggiornato il menu horizontal_menu.PHP, per inserire il link. La Figura 15 mostra la porzione di codice attraverso cui Viene creata una voce di menu(racchiusa tra i tag HTML), con un link (<a href> in notazione HTML) al percorso con cui si accede alla pagina news. La funzione site_url aggiunge il testo "/news" alla url, cioè www.pell.enea.it, trasformandola in www.pell.enea.it/news.

```
$route['attivazione-cittadino']     = 'enea/welcome/attivazione_cittadino';
$route['privacy-policy']            = 'enea/welcome/privacy_policy';
$route['news']                      = 'enea/welcome/news';
$route['background']                    = 'enea/welcome/background';
```

Figura 14-definizione dei path delle pagine in routes.PHP.

```
<li <?php if(site_url('/news') === current_url()) echo 'class="active"' ?>>
        <a href="<?php echo site_url('/news')?>">
                News
        </a>
</li>
```

Figura 15-aggiunta del link nell' horizontal_menu.PHP.

2.5 *Inserimento voci di menu nel menu Utente*

La presa in carico della gestione del sistema prevede anche l'eventuale aggiornamento di voci di menu, quando necessario.

In particolare, l'aggiornamento del menu utente prevede l'inserimento della voce di menu direttamente sul database.
La Figura 16 mostra il menu utente. In fase di inserimento di una voce, ed in base ai permessi dell'utente, è possibile visualizzare o meno alcune di esse quando viene effettuato il login, in modo che ogni gruppo di utenze visualizzi solo i link per i quali possiedono i permessi appositi. Alcuni link sono disponibili all'utente non autenticato.

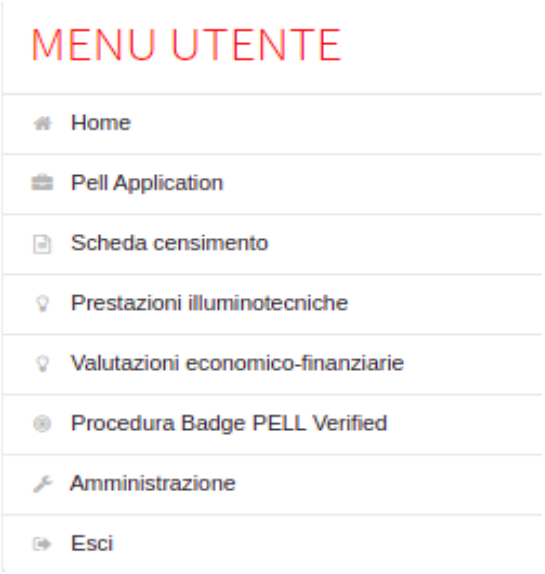

Figura 16-Voci di menu nel "menu utente".

2.6 Test sulle schede censimento

In seguito a tentativi di caricamento di schede censimento, è possibile che vi siano scarti da parte del sistema, dovuto a vari motivi.
Il sistema può scartare una scheda censimento perché:

- Il suo formato può non essere coerente con l'XSD (XML Schema Definition) che contiene i dati di validazione. Infatti la scheda censimento viene importata in formato XML, e necessita di una validazione formale tramite confronto con un file XSD .
- Può fallire la validazione. Oltre alla validazione tramite XSD, esiste una validazione tramite Schematron, un linguaggio che permette di validare i dati di un XML attraverso regole definite in un file.
- Possono esserci motivi sistemistici, dovuti alla struttura o configurazione hardware o software del server. Ne sono un esempio i limiti di memoria, o del numero di upload contemporanei, o ancora un disco fisso con poco spazio libero.
- Possono esserci anomalie nell'inserimento sul database o durante l'elaborazione. In alcuni casi, seppur in caso di validazioni XSD o SCHEMATRON con esito positivo, è possibile che vi siano incongruenze tra i dati da inserire nel database, e i campi della tabella che devono accoglierli. Ad esempio un dato può essere in formato testo (VARCHAR), ma il database si aspetta invece un numero (INTEGER)

L'elenco delle motivazioni di scarto di una scheda ovviamente non è esaustivo, ma nella maggior parte dei casi si rientra in una delle motivazioni indicate.
Nessun test può prevedere tutte le casistiche ma è importante prevenire e cercare di risolvere tempestivamente eventuali problemi rilevati dagli utenti finali.
In particolare, due schede hanno evidenziato dei problemi durante la sottomissione o durante il caricamento, richiedendo:

- La verifica della problematica
- L'incremento di parametri che limitavano la lavorazione delle schede

Nel primo caso la scheda non passava la sottomissione in quanto vi erano problemi di scadenza della sessione e di tempo limite di esecuzione dello script PHP che terminava la sua esecuzione non procedendo nella funzionalità. La scheda aveva una dimensione di circa 36 MB, pertanto il caricamento avveniva senza problemi, ma la durata prolungata della sottomissione, e quindi l'esecuzione della funzione di caricamento, terminava in modo improvviso.
La funzione che si occupa del caricamento delle schede, infatti, è scritta in PHP, e viene richiamata dal client attraverso una richiesta http POST che avviene tramite una chiamata Ajax (Asynchronous javascript and XML, un tipo di chiamata asincrona, che permette di scambiare dati tra il client e il server senza dover ricaricare la pagina); il controller elabora la richiesta e restituisce una risposta al client che l'ha effettuata.

In questo contesto, l'interruzione improvvisa del processo è avvenuta:

- A causa della durata di aggiornamento della sessione di CodeIgniter
- A causa di una errata configurazione del tempo di esecuzione degli script

Il primo problema è legato a CodeIgniter, infatti il tempo di sessione è un settaggio configurabile direttamente da codice, nel file config.PHP ed è indicato come **session_time_to update**, che insieme al **sess_expiration**, limita i tempi dell'esecuzione degli script legati alla sessione utente. Entrambi i valori sono stati incrementati.
La Figura 17 mostra la parte del file config.PHP in cui sono presenti i dati di sessione di CodeIgniter (in secondi) che sono i tempi di scadenza e di aggiornamento della sessione, il percorso di salvataggio nonché alcune opzioni di sicurezza. Se questi valori sono troppo bassi, l'applicazione potrebbe non riuscire a portare a termine l'esecuzione del programma.

```
$config['sess_expiration'] = 18000;
$config['sess_save_path'] = 'ci_sessions';
$config['sess_match_ip'] = FALSE;
$config['sess_time_to_update'] = 18000;
$config['sess_regenerate_destroy'] = FALSE;
```

Figura 17-Dettaglio del file config.PHP relativo ai parametri di sessione.

Il secondo problema è legato ad una errata configurazione di PHP (Hypertext preprocessor), un software installato sul server che permette di lavorare con il linguaggio PHP.
È stato dunque necessario incrementare il **max_execution_time** nel file PHP.ini(il file di configurazione presente nella directory del software PHP), cioè la durata massima, in secondi, di esecuzione di uno script PHP prima che venga interrotto. Il set di questo valore è molto importante per una applicazione, in quanto evita che script particolarmente lunghi ed elaborati possano bloccare il sistema.
Quando lo script viene invocato da interfaccia, di default è 30 secondi, mentre è illimitato per chiamate da riga di comando.
Nel secondo caso il problema da cui siamo partiti era la sua dimensione fisica, infatti si trattava di una scheda di circa 200 MB, decisamente fuori lo standard previsto da un caricamento sul portale.
Un caricamento di grandi dimensioni, in generale, genera varie problematiche che il sistema deve essere in grado di gestire:

- La fase di caricamento e acquisizione.
- La fase di decodifica dell'XML e creazione delle proprietà da salvare sul database.
- La fase di salvataggio vera e propria.

La prima fase è soggetta a vari problemi di tipo sistemistico, infatti come visto in precedenza, ci sono dei limiti fisici di caricamento su un server, dichiarati nella configurazione di PHP.

Questo limite è inserito di default per evitare che uno o più files di grandi dimensioni possano bloccare il sistema per molto tempo tuttavia, al fine di caricare documenti di dimensioni maggiori, può essere incrementato secondo necessità.
Quando viene implementato un metodo particolarmente impegnativo per un sistema (ad esempio che utilizza molta memoria oppure inserisce molti dati in un database, oppure necessita di molto tempo per la sua esecuzione), potrebbero esserci gravi problemi di performances che alterano la funzionalità di una web application. In generale questi task sono gestiti attraverso funzionalità che lavorano in background e che siano quindi slegati dalla sessione utente. Infatti la lavorazione di una pratica può generare problemi bloccando l'utente su una pagina di caricamento, impossibilitato a chiudere il browser per evitare l'eventuale perdita di messaggi di errori a front end.
Al fine di caricare la scheda, è stato dunque necessario incrementare il valore indicato nel campo **max_upload_size** , che rappresenta la dimensione massima di un file di cui effettuare l'upload, e che deve essere proporzionato alle dimensione dei files da caricare sul sistema. Di default il valore era impostato a 128 MB, insufficiente per le nostre necessità. Tale problematica era comunque visibile nei logs di PHP. La Figura 18 mostra gli errori nel file dei logs di Apache, generati dal fallimento nel caricamento della scheda censimento. I logs sono tutte quelle informazioni di info, warnig o error che possono essere scritte al fine di identificare meglio eventuali problemi o di controllare l'andamento dei vari processi. In particolare nel nostro caso, l'application server Apache ha generato degli errori molto parlanti che sono stati inseriti in formato testo nel suo file di log, rendendoli così disponibili.

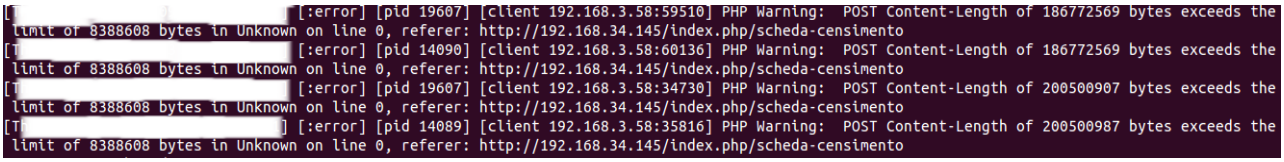

Figura 18-Dettaglio degli errori nel file dei logs di Apache.

Tuttavia l'incremento di questo valore non è bastato alla risoluzione del problema, infatti, sebbene l'upload andasse a buon fine, avveniva comunque l'interruzione improvvisa dello script di inserimento. Si è pertanto proseguita l'analisi.
Un altro errore segnalato nei logs è stato un java.lang.OutOfMemoryError: Java heap space, un tipico errore che indica che lo script PHP necessita di più memoria rispetto a quella permessa dalla configurazione di PHP . La Figura 19 mostra il dettaglio del file di log, in cui è stata anche riportata l'eccezione dovuta ad un problema di memory limit per lo script PHP relatio al caricamento della scheda censimento. Infatti qualora il processo lanciato dallo script richiedesse un uso troppo elevato della memoria, esso non potrebbe procedere se non in seguito ad un incremento del valore limite della stessa.

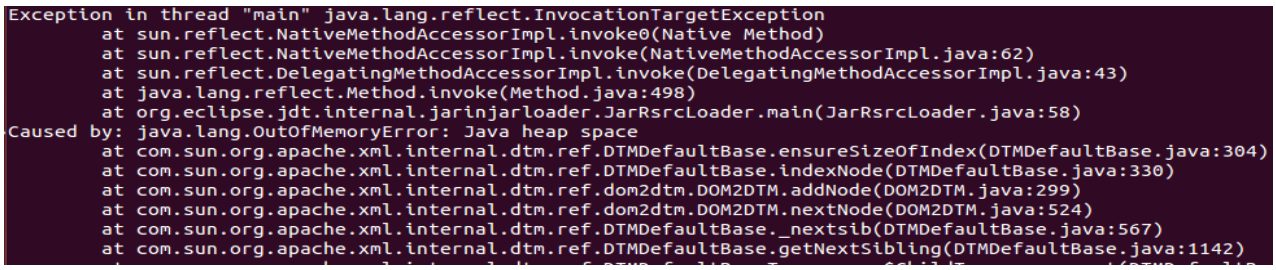

Figura 19-dettaglio del log relativo all'eccezione dovuta ad un problema di memory limit per lo script PHP relativo al caricamento della scheda censimento.

Al fine di risolvere il problema, è stato aumentato il parametro di PHP, definito **memory_limit,** che setta il massimo quantitativo di memoria in byte che uno script può allocare per la sua esecuzione. Per default, questo valore viene settato a 16 MB, valore decisamente troppo basso per le nostre esigenze.

In seguito al suo incremento a 128 MB e successivamente prima a 256 MB e poi a 512 MB, l'errore nei log è scomparso, tuttavia lo script continuava ad interrompersi sempre nello stesso punto.
Questo ha portato ad analizzare varie altre possibilità, come un problema legato all'inserimento dei dati sul database, per cui sono stati analizzati i log di MySQL senza alcun risultato, o problemi di sessione come per la scheda di La Spezia, o ancora problemi legati al keepalive della sessione TCP o di Apache.
L'assenza di logs ha reso complessa l'analisi, pertanto abbiamo provato a "estrarre" il metodo di upload e inserimento, slegandolo dalla sessione e creando un task in background che potesse importare la scheda eliminando tutto ciò che riguardava la richiesta http chiamando il metodo da riga di comando e quindi bypassando le limitazioni imposte dalla sessione e dal web server.
A tal fine è stata utilizzata una libreria, scaricata da Github (un repository open source per la gestione e il versioning di applicazioni), denominata "BackgroundProcess.PHP" (Background Processer) che ci ha aiutato nello scopo.
È stato creato quindi un controller Background.PHP, che estende il MY_Controller, in cui è stata riportata la funzione upload_post() di caricamento della scheda censimento, e che attualmente è in un controller di tipo REST (representational state transfer, un modello architetturale di condivisione dati basato su HTTP e JSON) per gestire la POST http.
La nuova funzione public run () prende la scheda censimento già caricata in una cartella "uploads" ed esegue la decodifica dell'XML relativo alla scheda censimento, la validazione della scheda tramite XSD (l'XSD è un file che contiene alcune regole che un file XML deve rispettare), e l'inserimento sul database sfruttando le potenzialità di CodeIgniter e del model già esistente.
La funzione viene chiamata dal metodo descritto nella Figura 20, che essenzialmente rappresenta una chiamata attraverso cURL (client URL), un software open source che si utilizza tramite riga di comando per comunicare direttamente con un server invece di passare attraverso un browser. Viene quindi invocato il metodo run(), implementato nel nuovo controller, passando dei parametri relativi alla scheda censimento e necessari al suo caricamento.

```
$this->load->library('backgroundProcess');
$this->backgroundprocess->setCmd("curl -o /home/mg/Scrivania/log/log_background_process.log " .
        base_url('/background/run/'.$label.'/'.$schede_current.'/'.$schede_type.'/'.$file_type.'/'.$file_name.'/'.$raw_name);
$this->backgroundprocess->start(false);
```

Figura 20-La figura mostra la modalità con cui vengono invocate le funzioni di BackgroundProcess e cURL per lanciare il metodo di caricamento della scheda, attraverso il metodo run(), da riga di comando.

In seguito al test di questa funzionalità, e alla sua esecuzione nell'ambiente di test, il caricamento non è andato a buon fine, ma il processo da background ha segnato nei log un errore durante il processo dovuto ad una funzione specifica di CodeIgniter che andava in errore a causa di limiti di memoria durante la sua esecuzione.
La funzione in errore, è una funzione relativa ad una libreria di sistema, in particolare mysqli_result.PHP che è disponibile in CodeIgniter nella cartella system/database/drivers/mysqli.
Pertanto abbiamo provveduto ad incrementare ulteriormente il **memory_limit** fino a 1024 MB che ha finalmente permesso di risolvere il problema di caricamento.
Al fine di minimizzare le probabilità di incorrere in problemi simili, sono stati incrementati anche altri parametri di configurazione nel file PHP.ini come ad esempio il numero di variabili di input necessarie per le richieste che possono essere accettate (max_input_vars) oppure il massimo tempo in secondi che uno script necessita per decodificare i dati legati alla richiesta di GET o POST (max_input_time).

Altri test sulle schede censimento sono stati fatti in seguito alla modifica del flusso di gestione delle stesse, ad esempio l'invalidazione della scheda in fase di modifica di una scheda validata, al fine di testarne nuovamente la validità, oppure la sottomissione.
Oltre ai test e il sostegno alle attività di sviluppo, il PELL IP prevede una parte di infrastruttura dedicata all'acquisizione, salvataggio e monitoraggio dei dati di consumo provenienti dagli "smart meters", dei dispositivi intelligenti che registrano i dati elettrici e comunicano direttamente con il fornitore.

Questa parte prende il nome di “fase dinamica”, e ci ha visti partecipi come supporto allo studio dell’infrastruttura dedicata ad accogliere ed elaborare i dati, ma soprattutto all’implementazione di particolari e specifici KPI (Key performance index).

3 PELL edifici

Il valore di un portale come quello del PELL IP, consiste nella effettiva possibilità di censire, e quindi studiare, analizzare e valutare, un intero impianto elettrico (dal POD, alla localizzazione geografica, al singolo punto luce) situato in un comune. La conoscenza nel dettaglio di un impianto di Illuminazione pubblica, è molto importante per la sua riqualificazione, per il risparmio economico ed energetico, e per la semplificazione dei processi di manutenzione ordinaria e straordinaria dello stesso.
Il PELL Edifici nasce con la stessa idea: quella di censire gli edifici pubblici, in termini architettonici ed energetici, per rendere le amministrazioni consapevoli del patrimonio immobiliare pubblico, promuovendone anche la riorganizzazione e la riqualificazione.
Molti sono gli edifici pubblici che possono essere censiti e monitorati, dalle sedi istituzionali alle scuole, ma anche, e soprattutto, edifici maggiormente energivori come, ad esempio, gli ospedali.
La complessità dell'analisi è dovuta principalmente alle diverse tipologie di edifici da censire; infatti la struttura di una scuola è diversa da quella di un ospedale o di una caserma, come diverse sono le informazioni peculiari relative all'immobile.
Per questo motivo, l'infrastruttura del nuovo portale deve tenere conto di queste differenze di ambiti, dando la possibilità di poter gestire varie tipologie di edifici, attraverso la definizione del tipo di informazioni che sono necessarie a descrivere i loro diversi comportamenti.
Si è partiti dalle scuole, un grande patrimonio immobiliare italiano, come infrastruttura da censire e monitorare.
Analogamente al PELL IP è stata predisposta una scheda censimento per raccogliere le informazioni sugli edifici, al fine di caratterizzarli anche da un punto di vista energetico, cercando di rendere possibile un sistema di "diagnostica" che permetta di rilevare inefficienze e al contempo fornire soluzioni.
Ovviamente, l'ambito dell'illuminazione pubblica è notevolmente differente dall'ambito residenziale, pertanto è necessario creare una nuova infrastruttura adatta a gestire quest'ultimo, e questo può avvenire attraverso lo sviluppo di un nuovo portale dedicato.
In questo ambito, l'attività dell'università Roma Tre è stata quella di essere di supporto nella scelta della struttura dati migliore per l'implementazione del nuovo Portale, anche sulla base della nostra pregressa esperienza dovuta allo studio approfondito del PELL IP e delle sue problematiche.
Pertanto abbiamo effettuato attività di consulenza nella scelta progettuale del sistema di persistenza effettuando un confronto tra varie tipologie di database Relazionali(MySQL) e NoSql(MongoDB, ElasticSearch), al fine di valutare lo strumento di storicizzazione che possa fornire il miglior compromesso tra performance ed efficienza, evidenziando come, per la struttura dati oggetto dell'analisi, sia auspicabile l'utilizzo di un database NoSql ed in particolare ElasticSearch.

3.1 Infrastruttura dati

Da un punto di vista tecnico, è sempre conveniente strutturare un portale che sia in qualche modo "estendibile" anche a varie tipologie di edifici, non solo le scuole, in modo da ottimizzare le risorse e lavorare su parti "comuni" (come ad esempio interfacce grafiche di visualizzazione/gestione, indipendenti dal tipo di edificio che si vuole censire) e parti custom (le schede censimento che invece sono specifiche per ciascun edificio).
L'esperienza del PELL IP, sviluppato attraverso CodeIgniter su un database MySQL (vedi paragrafo 2.4), è stata di fondamentale importanza per lo studio e l'analisi di un nuovo portale, nei confronti del quale abbiamo fornito assistenza alla sua progettazione.
Infatti, anche se l'illuminazione pubblica ed il residenziale siano ambiti profondamente differenti, da un punto di vista architetturale e della gestione dei dati cambia molto poco.
Le problematiche relative al PELL IP, descritte anche in parte nei paragrafi precedenti (es. paragrafo 2.6), hanno mostrato una particolare debolezza proprio per quanto riguarda l'infrastruttura di storicizzazione dei dati relativi alle schede censimento, cioè il database.

Il PELL IP si basa sull'utilizzo di un RDBMS (relational database management system) MYSQL, in cui la scheda censimento viene salvata.
La Figura 21 mostra il diagramma ER (Entity relationship) relativo alla struttura della scheda censimento, che ricalca la struttura gerarchica dell'XML. Si può notare la complessità della struttura e la grande quantità di informazioni sull'impianto che vengono salvate.

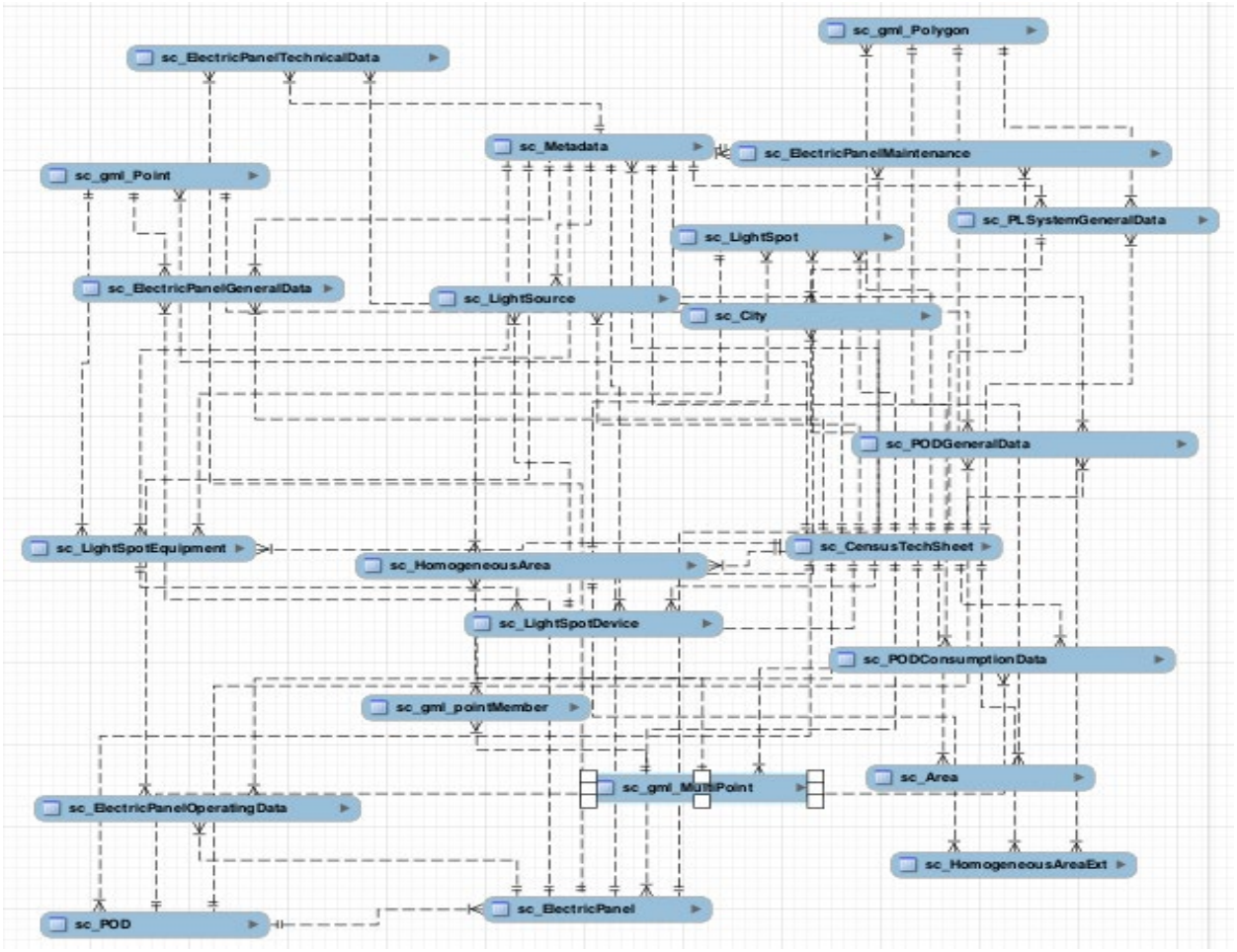

Figura 21-ER diagram (diagramma delle relazioni tra le entità, cioè le tabelle) delle tabelle relative alla Scheda censimento.

Questo diagramma evidenzia quali sono le tabelle che mappano la scheda censimento (ad esempio sc_CensusTechSheet, oppure sc_POD ecc..) e, attraverso frecce e link, le relazioni tra le stesse. Ogni tabella è interessata in un modo specifico nel processo di caricamento/visualizzazione/aggiornamento di una scheda censimento, e contiene le informazioni presenti nell'XML eventualmente modificate e/o elaborate.
La struttura della scheda censimento, a livello di database, come si può vedere dalla figura è abbastanza complessa ed interessa circa 23 tabelle sul database.
Le tabelle, cioè le entità fisiche che contengono i dati relativi all'entità logica, hanno chiavi primarie ed esterne per relazionarsi ad altre tabelle. Queste "chiavi" sono dei vincoli che identificano univocamente il singolo record della tabella (chiave primaria), oppure le relazioni tra varie tabelle (chiave esterna).
Ogni tabella è costituita da campi, ciascuno contenente un tipo di dato.
Solo a titolo di esempio si riporta in Figura 22, la tabella sc_CensusTechSheet che contiene i dati generali della scheda censimento. Si evidenziano in particolare il nome della scheda, l'id, il tipo di scheda (ante/post riqualificazione), se è l'ultima scheda inserita per l'impianto, se è in stato congelato, oppure malformata, oppure invalida, e le date relative alla creazione e/o aggiornamento.

id	label	SchedeType	SchedeCurrent	user_id	submitted	is_hibernated	bad_formed	is_valid	start_hibernation	namespaces	created_at	updated_at
182	census-sheet-LASPEZIA-v581-rev04.xml	1	1		2021-02-26 18:04:14	1	0	1	365	ns1,xsi,ns2	2021-02-10 09:24:24	
164	PELL_GENOVA_red_20210129123441	1	false	1		0	0	1	365	ns1,xsi,ns2	2021-01-29 12:38:27	
143	SchedaCensimentoEsempioNonUniforme_20	2	true	1	2021-02-18 10:55:22	0	1	1	365	gml,xsi,xlink	2020-11-24 12:18:07	2021-02-18 10:53:46
142	SchedaCensimentoEsempioNonUniforme_20	1	false	1	2021-02-18 11:06:37	0	1	1	365	gml,xsi,xlink	2020-11-24 12:16:25	
140	SchedaCensimentoEsempio2_20200720093	4	false	1	2021-02-18 11:38:08	0	1	1	365	gml,xsi,xlink	2020-07-20 09:38:07	2021-02-17 21:19:06
139	SchedaCensimentoEsempio1_20200720092	1	false	1		0	0	0	365	gml,xsi,xlink	2020-07-20 09:28:07	2021-02-17 21:15:17

Figura 22-La tabella relativa ai dati della scheda censimento.

I dati vengono mostrati sull'interfaccia del PELL, come visibile nelle Figure 23 (che mostra alcuni dati salvati nella tabella sc_CensusTechSheet, ed in particolare il nome della scheda, l'id, il tipo di scheda, se è una scheda corrente e il suo stato: malformata o invalida, segnalati graficamente nel campo "Scheda corrente", e le date di aggiornamento e sottomissione) e 24 (che mostra alcuni dati anche all'interno della scheda censimento).

Figura 23-Elenco tabellare delle schede censimento. Per ogni record sono mostrati alcuni dati salvati nella tabella sc_CensusTechSheet

Figura 24-Alcuni dati della tabella sc_CensusTechSheet vengono riportati all'interno dell'anagrafica della singola scheda.

Le altre tabelle corrispondono alle varie sezioni della scheda censimento, accessibili dall'interfaccia del PELL.

Quando viene invocata da interfaccia la funzione di caricamento di una scheda XML, la funzione, oltre a validare la scheda, elabora i dati presenti nel file XML, li modifica secondo quanto richiesto dalle specifiche, e infine crea delle "INSERT", cioè delle query di inserimento sulle varie tabelle, oppure degli "UPDATE", cioè delle query di aggiornamento dei dati già salvati sul database.
Oltre a creare queste query di inserimento e aggiornamento, è altresì possibile che durante questa fase sia necessario richiamare alcune informazioni già memorizzate nel database, pertanto è possibile che vengano effettuate anche delle query di ricerca ("SELECT") per il recupero dei dati.
È importante, in un database relazionale, che le query siano generate in una sequenza corretta, altrimenti, quando vengono eseguite in una sequenza errata, possono portare a problemi di consistenza e coerenza di dati, facendo fallire l'importazione o, peggio, salvando i dati in modo non corretto.
Normalmente, viene definita all'interno della funzione una "Transazione", cioè una sequenza di operazioni sul database che, in caso di fallimento, riporta il database in uno stato di coerenza, effettuando il salvataggio effettivo (definito COMMIT), solo alla fine della transazione. Durante la transazione vengono effettivamente compiute tutte le operazioni di SELECT, INSERT e UPDATE, ma solo il COMMIT salva in modo definitivo le modifiche al database.
Poiché ogni operazione sul database interessa generalmente una tabella, e vengono effettuate una alla volta, in caso di grandi quantità di dati, e cioè di molte operazioni consecutive sul database, i tempi di esecuzione risultano dilatati, necessitando un impegno del database notevole in termini di performances.

Una volta effettuato il COMMIT, cioè una volta salvate definitivamente le modifiche al database, l'importazione è conclusa.
Anche quando il database viene interrogato in fase di ricerca di informazioni, ad esempio quando si apre la scheda censimento da interfaccia web, si deve prevedere una particolare attenzione alla definizione di alcuni indici (delle strutture che dovrebbero velocizzare la ricerca dei dati in un database in quanto permettono di non cercare l'informazione che si desidera su tutti i dati presenti, ma solo su un sottoinsieme caratterizzato da uno o più indici) per evitare che i tempi di risposta del database siano troppo lenti.
Da questa breve introduzione sui database relazionali, e da quanto già descritto nel paragrafo 2.6, relativamente all'importazione di grandi schede censimento, si possono evincere i limiti di questa struttura che, all'aumentare del numero dei dati, genera:

- Lentezza nel processo di salvataggio dei dati.
- Difficoltà nella visualizzazione a schermo dei dati salvati.

Un altro limite dei databases relazionali è che ad ogni modifica della struttura dati, ad esempio dovuta all'aggiunta di un campo sul XML della scheda censimento, deve corrispondere una modifica della struttura del database, inserendo manualmente la colonna nella tabella relativa.
Pertanto, sebbene i normali database relazionali siano pensati per gestire e storicizzare grandi quantità di informazioni, le fasi di salvataggio dei dati e di recupero degli stessi sono molto delicate e soggette a varie problematiche, e ciò comporta che il loro utilizzo e la loro configurazione debbano essere pensati in base all'applicazione che deve essere sviluppata.

I limiti di un database relazionale come MySql, possono in parte essere superati dai database NoSQL, cioè strutture dati che non utilizzano entità relazionali bensì documenti che vengono salvati direttamente in un formato specifico e dipendente dal database utilizzato.
Queste strutture non salvano i dati su tabelle legate tra loro, ma salvano l'oggetto in un documento che può contenere qualsiasi tipo di informazione.
Si può quindi intuire che, nei casi da noi considerati, quindi di schede censimento di edifici o impianti di illuminazione pubblica, il salvataggio delle stesse non avverrebbe creando delle operazioni come descritto, bensì scrivendo i dati della scheda su un documento e salvando direttamente quest'ultimo.
Il punto forte di questi sistemi sono, ovviamente, le performances e la flessibilità. Infatti, a fronte in alcuni casi di una potenziale maggior complessità nel richiamare i dati, in quanto le varie operazioni su questi database prevedono sintassi specifiche e non sempre facili da descrivere, i tempi di inserimento e ricerca sono notevolmente ridotti rispetto ai database relazionali.
Ad esempio, attualmente l'inserimento di una scheda censimento contenente 50.000 punti luce prevede circa 60-90 minuti, mentre utilizzando un database NoSql l'inserimento è quasi istantaneo.
Inoltre la possibilità di salvare un documento elimina il vincolo di avere una struttura dati fissa pertanto, se la scheda censimento XML dovesse prevedere l'aggiunta di un nuovo campo, il database NoSql potrebbe non necessitare di alcuna modifica.
La figura 25 riassume in breve le differenze tra i due tipi di database, relazionale e NoSql.
Ormai sono disponibili vari database NoSql, e la nostra attenzione si è focalizzata in particolare su MongoDB ed ElasticSearch, valutando poi quest'ultimo come scelta definitiva per la storicizzazione dei dati relativi al nuovo portale PELL EDIFICI.

È in corso di valutazione anche la migrazione del portale PELL IP su un database ElasticSearch.

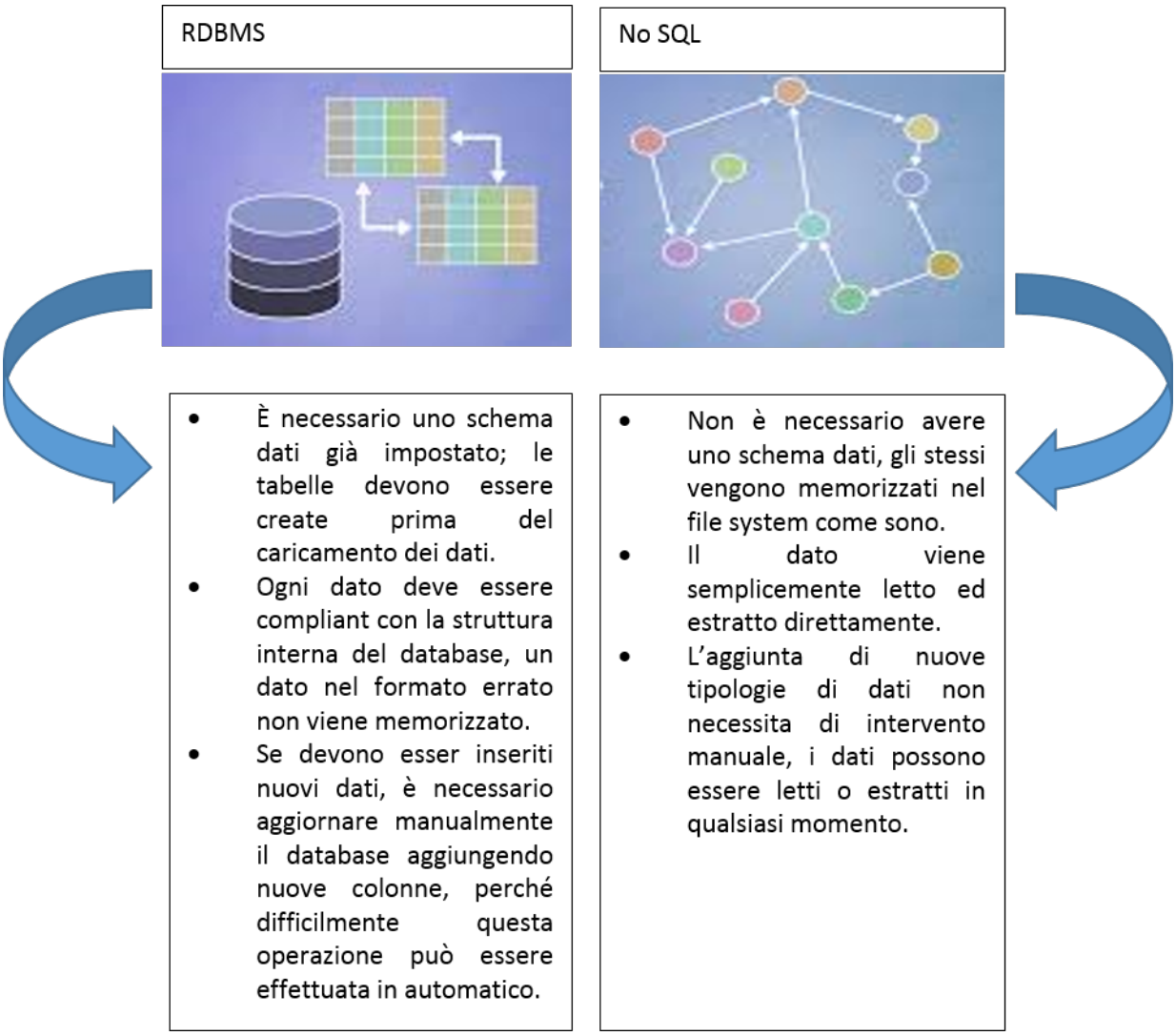

Figura 25 – La figura evidenzia le differenze sostanziali tra un database NoSql e un database relazionale.

3.2 Caratteristiche di ElasticSearch

Oltre alle caratteristiche proprie dei database NoSQL, ElasticSearch presenta altre peculiarità che ci hanno fatto valutare la sua applicabilità alla struttura che stiamo implementando.
ElasticSearch è essenzialmente un motore di ricerca open source in grado di gestire tutti i tipi di dati, e gestisce dati strutturati e non. Questo vuol dire che è molto flessibile e che offre una elevata possibilità di customizzazione in base alle necessità.
È una applicazione scalabile, cioè può gestire anche i BigData senza problemi, ottimizzandone la storicizzazione e la richiesta.
I documenti sono salvati in formato JSON, sono raccolti in "Index" e ciascuno appartiene ad un "Type", e possiede un "id" univoco ed alcune proprietà.
Un indice può contenere anche vari tipi di documenti.

La Figura 26 mostra la struttura di un cluster ElasticSearch.

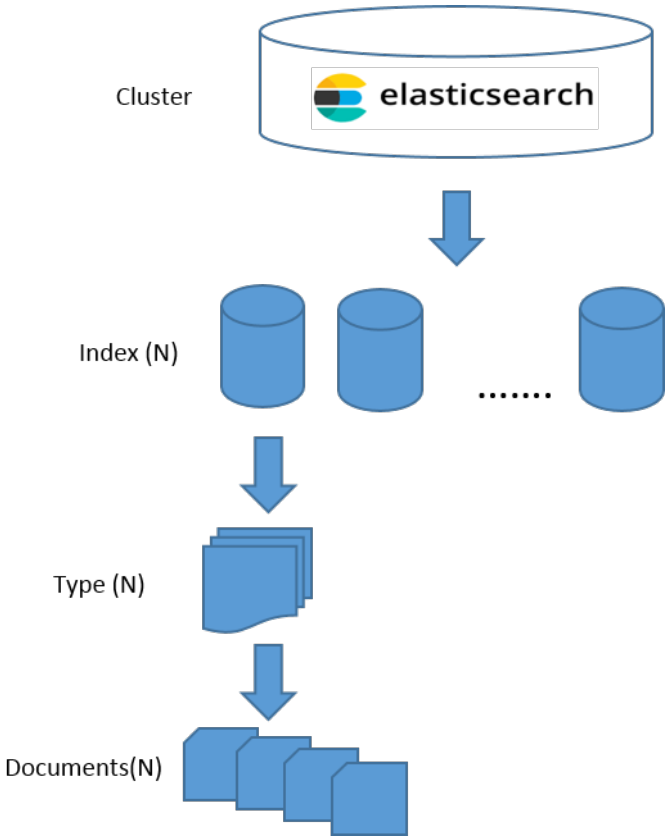

Figura 26- Un cluster ElasticSearch contiene più Indexes; ciascun Index contiene più Types; ciascun Type contiene molti Documents, ciascuno dei quali ha delle proprietà.

Le operazioni sui Documents sono effettuate attraverso delle API REST che sono esposte e, quindi, disponibili alle richieste.
Le APIs REST di ElasticSearch (http://elastic.co/guide/en/elasticsearch/reference/current/docs.html) sono funzioni facilmente invocabili tramite una richiesta http e permettono di effettuare vari tipi di operazione come inserimento, modifica, recupero dati.
Le API possono essere invocate in vari modi, ad esempio passando le query nell'URL, e quindi all'interno della chiamata, oppure le query possono essere espresse in formato JSON (Query DSL).
Esistono varie funzioni già definite, ad esempio la ricerca di tutti i documenti, oppure una ricerca diretta per termine o per intervallo, o ancora, è possibile comporre query più complesse.
La scelta di ElasticSearch sicuramente rappresenta un punto di svolta rispetto a ciò che è attualmente in essere nel PELL IP, presentando tutte le caratteristiche che permettono di sviluppare una applicazione efficiente e sicura.

4 Gestione notifiche tramite Check_MK

La gestione dei vari progetti in ambito Smart City, prevede l'utilizzo di diversi server che forniscono servizi, piattaforme o processi.

Al momento l'ambito Smart City di ENEA comprende circa 20 server, di cui alcuni ospitano la piattaforma Big Data comprendente Hadoop e Spark, altri ospitano i portali web PELL IP di test e produzione (sia front end che database), altri ancora gestiscono i servizi (MQTT) di interazione tra Hadoop e PELL o Hadoop e i dati provenienti dagli smart meters ecc. Molti server interagiscono tra loro scambiandosi informazioni, dati e processi.

Questo evidenzia una complessità dell'infrastruttura che rende necessario un monitoraggio della stessa che garantisca che tutto funzioni correttamente e, qualora ciò non dovesse avvenire, deve tempestivamente essere messo in atto un processo di gestione del problema.

Più complessa è l'infrastruttura, maggiore è la possibilità che si verifichino situazioni impreviste, e particolarmente difficoltosa risulta l'identificazione del problema, che passa attraverso la verifica manuale da parte di un sistemista, di ogni server impattato.

Al fine di gestire questa infrastruttura complessa, attraverso il controllo puntuale del funzionamento di ogni server appartenente ad essa, è stato implementato un sistema di monitoraggio che fornisca informazioni sullo stato dei sistemi, nonché dei singoli servizi o processi di interesse.

Un sistema di monitoraggio è rivolto a due tipologie di utenti: il tecnico sistemista che si occupa materialmente della gestione dei server e dei processi ospitati su di essi, e il generico utente che vuole controllare e avere sotto controllo lo stato dei sistemi a scopo informativo.

Esistono vari sistemi di monitoraggio che offrono le più disparate modalità di gestione dei server e dei processi, ma nel nostro caso è stato scelto un software open source customizzabile in base alle necessità, totalmente gestibile da ENEA in autonomia, e che presenta una varia tipologia di plugins e tools per gestire le più svariate situazioni.

Nagios, con la sua soluzione open source basata su Check_MK, provvede a metodi e funzioni che permettono di controllare ogni aspetto del sistema, dai singoli servizi, alle applicazioni, all'attività degli hosts, ecc., inoltre la possibilità di gestire dati storici, e una interfaccia web da cui è possibile monitorare o configurare secondo le necessità, rendono questo strumento molto interessante.

Attraverso l'interfaccia di Check_MK è dunque possibile configurare check specifici per ogni server, oppure per gruppi di servers o ancora per singoli processi o gruppi di processi. Questa operazione di definizione e implementazione dei controlli nel sistema di monitoraggio, è stata effettuata da altre società, mentre noi come Università ci siamo occupati nel dettaglio di gestire le eventuali notifiche generate dal sistema di monitoraggio.

Infatti, se da un lato i check si occupano di identificare concretamente i problemi, le notifiche di questi ultimi forniscono la visibilità necessaria alla loro gestione tempestiva, sia ai tecnici che devono risolvere il problema, sia al generico utente che vuole essere informato dello status dei sistemi.

In quest'ottica Check_MK permette la generazione di messaggi (le notifiche), sulla base dei controlli implementati, e il successivo invio degli stessi verso uno o più utenti/gruppi di utenti, al fine di fornire informazioni sullo lo stato dei servizi attivi via email oppure sms.

Al fine di gestire le notifiche, è necessario:

- Configurare uno o più gruppi di utenti. Ad esempio si può creare un gruppo di utenti che ricevano notifiche solo dai check relativi ai portali WEB, oppure un gruppo che riceva notifiche dai servizi legati

all'infrastruttura Big Data, o ancora un gruppo di amministratori che necessitano di controllare tutti i server e i processi attivi ecc..

- Creare gli utenti e associarli ai gruppi.
- Creare delle regole di notifica in base a criteri scelti, e decidere cosa notificare e a quali utenti o gruppi. Infatti è possibile che un sistemista necessiti di informazioni di dettaglio relative al problema o anche dai dati di performance, mentre un utente normale può necessitare di conoscere solo lo stato di attività di alcuni sistemi.

Di seguito vengono descritte le modalità e gli step necessari alla configurazione del sistema di notifiche basato su Check_MK.

Esso prevede un tool chiamato WATO (Web Administration Tool) che prevede una interfaccia grafica (GUI-graphic user interface) per la gestione del sistema. WATO presenta un menu di amministrazione da cui è possibile accedere a tutte le funzionalità offerte dal sistema di monitoraggio.

4.1 Configurazione gruppi di utenti

Nella sezione WATO-Configuration del frontend di Check_MK, attraverso il link "contact group" è possibile definire gruppi di utenti, ad esempio:

- All: predefinito da Check_mk.
- Controllers: gruppo di utenti che ricevono le notifiche relative ai server in up o down, nonché i singoli servizi su tutti gli hosts, configurato da noi.

È possibile inserire nuovi gruppi per singolo host/gruppi di hosts.

La figura 27 mostra la sezione "WATO-Configuration" (sulla destra), da cui è possibile effettuare l'accesso alla sezione "Contact Groups" per la Configurazione di gruppi di utenti. Attraverso il tasto "New Contact group" è possibile definire gruppi di utenti, ad esempio controller, oppure gestori_portali ecc.

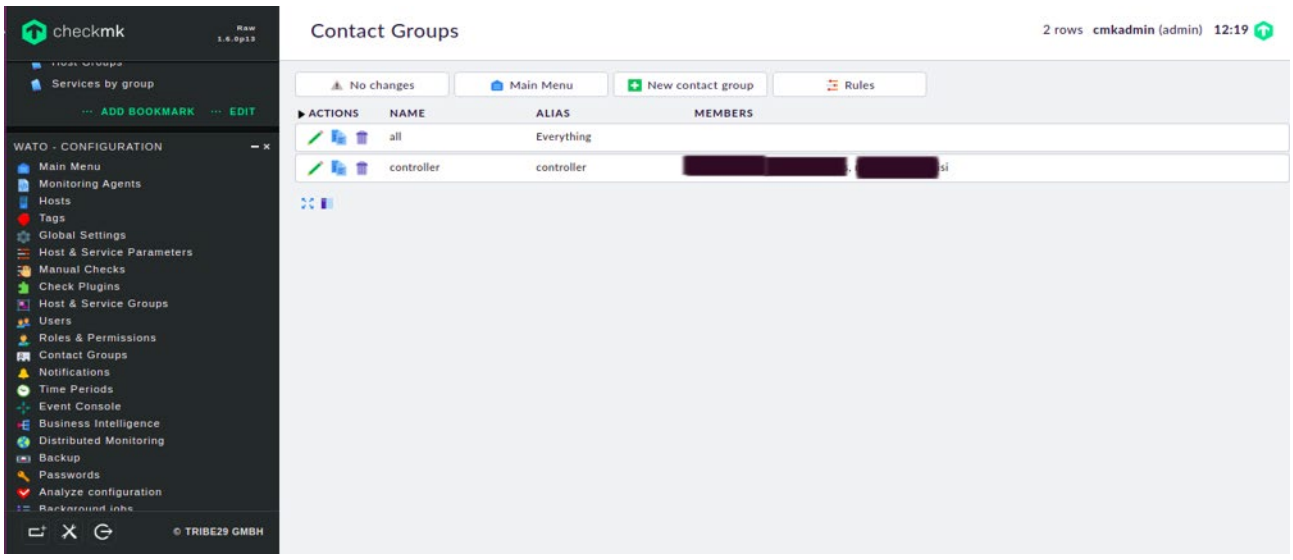

Figura 27- sezione "WATO-Configuration" da cui è possibile effettuare l'accesso alla sezione "Contact Groups" per la Configurazione di gruppi di utenti.

In seguito alla creazione di gruppi, devono essere inseriti gli utenti dal link WATO-Configuration – "Users".

In fase di creazione utente possono essere associati uno o più gruppi. È importante configurare un indirizzo email corretto, in quanto le notifiche sono spedite ad ogni indirizzo email dei contatti.

Attraverso il link “Users” nel menu “Wato-configuration” è possibile efettuare l’accesso alla sezione “Users”. Qui sono presenti tutti gli utenti già creati nel sistema con un riepilogo di informazioni sugli stessi e la possibilità di modificarli, eliminarli o aggiungerne di nuovi (vedi Figura 28).

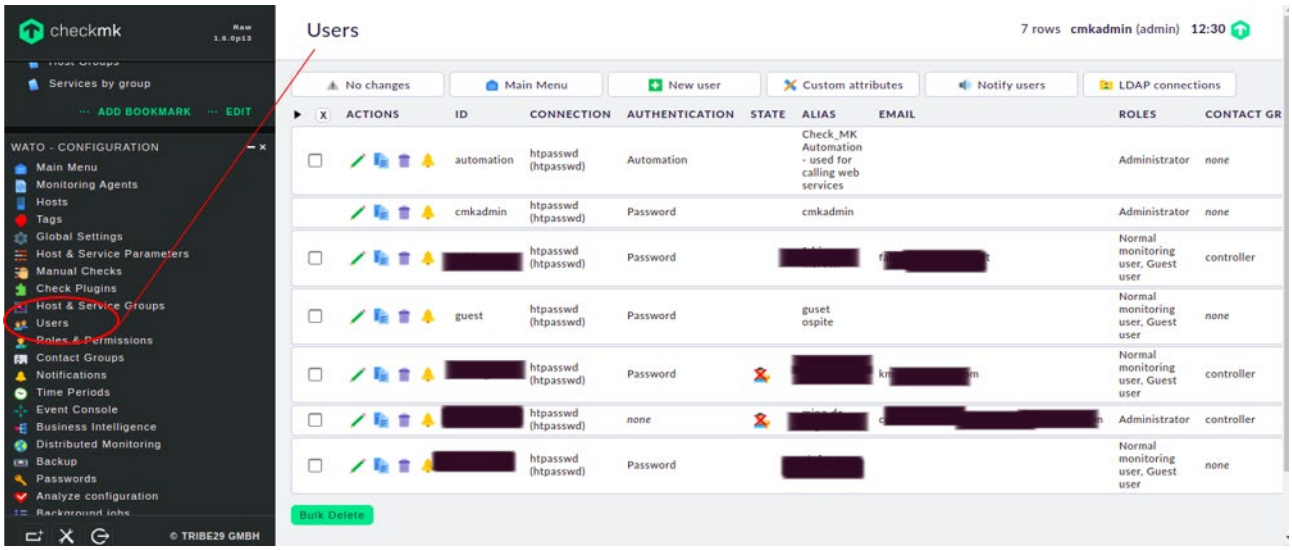

Figura 21-Attraverso il link “Users” nel menu “Wato-configuration” è possibile efettuare l’accesso alla sezione “Users”.

4.2 Creare regole di notifica.

La sezione “WATO-Notifications” permette la configurazione delle regole di notifica; il link “Notifications” permette l’accesso alla sezione omonima (Figura 29), in cui è visibile l’elenco delle regole di notifica create (Global notification rules), con le relative condizioni e contatti. Da qui è possibile accedere in edit alla notifica, oppure eliminarla.

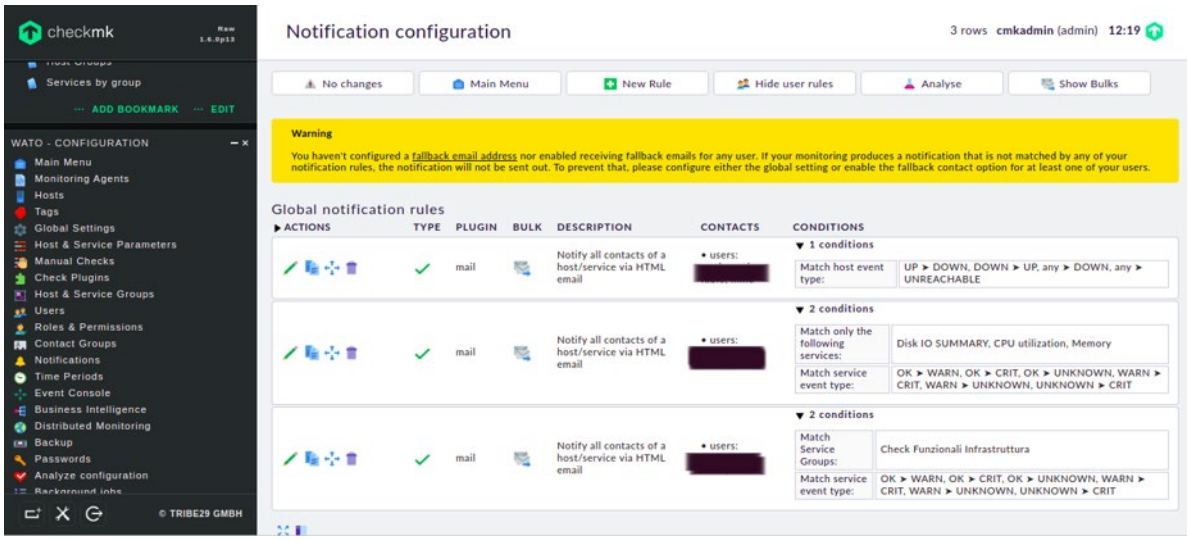

Figura 29-Il link “Notifications” permette l’accesso alla sezione omonima, in cui è visibile l’elenco delle regole di notific configurate.

La gestione delle notifiche da front-end è molto semplice; è anche possibile creare script di notifica custom, tuttavia, questa operazione potrebbe non essere necessaria vista la flessibilità del modulo WATO.

Ogni regola può essere customizzata in modo da:

- Definire a quali gruppi/utenti deve essere inviata l'email e se gli utenti possono disabilitare la regola.
- Customizzare il modello di email e i dettagli da inviare, compresa la possibilità di effettuare il bulking delle notifiche.
- Definire quali eventi generano una notifica, ad esempio eventi legati agli hosts e/o ai servizi.

Alcune regole che al momento possono interessare i sistemi attualmente monitorati, e che abbiamo provveduto a configurare, sono le seguenti:

- Per tutti gli hosts si verifica se la macchina è attiva (up), spenta (down), irraggiungibile (unreacheble).
- Per tutti gli hosts si verifica se gli specifici servizi relativi a RAM, hard disk e CPU risultano ok (esempi di servizi sono: CPU utilization, Disk IO SUMMARY, Memory).
- Per alcuni hosts può essere monitorata l'attività sulle singole porte (es WEB GUI), oppure su check specifici (es Namenode, Servizi).

Dalla sezione WATO-notification, è possibile accedere alla sezione relativa.

Di default, Check_MK prevede una regola di notifica che deve essere customizzata accedendovi dall'apposito link. La Figura 30 mostra un esempio di notifica in edit in cui è visibile la sezione "Rule Properties". In essa è possibile definire il nome della notifica, la sua attivazione e la possibilità per un utente di disattivarla al fine di non ricevere messaggi.

La sezione "Rule properties" presenta una descrizione e dei flags per:

- Inibire la sua applicazione
- La sua eventuale disattivazione da parte degli utenti.

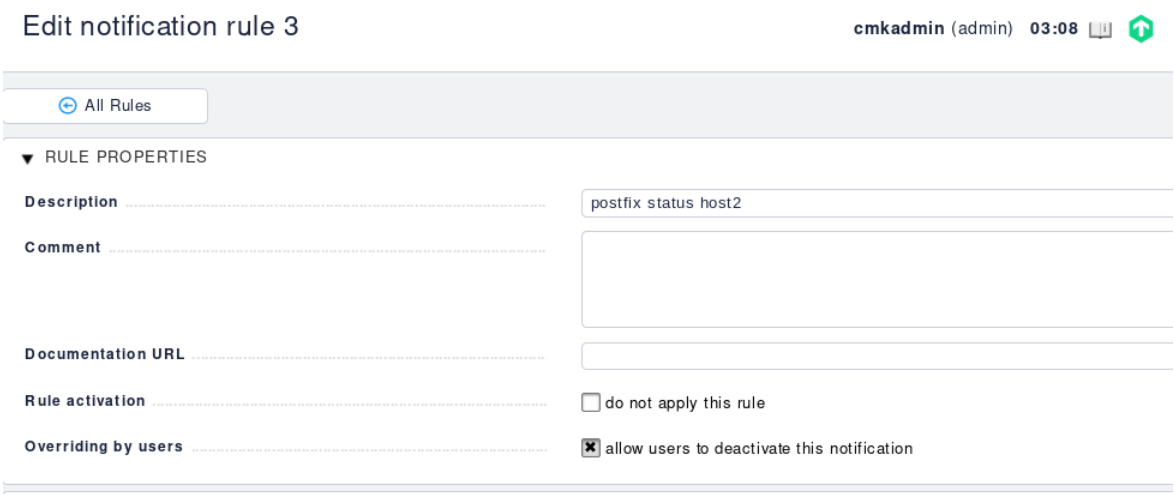

Figura 30- la sezione "Rule Properties", accessibile accendendo in edit nella regola di notifica

In Figura 31 è mostrata la sezione "Notification methods" di una regola di notifica, che definisce il metodo di notifica, cioè email, notifiche push, sms ecc.. Nel nostro caso abbiamo selezionato "Email HTML" per produrre una email con i parametri da definire di seguito, tra cui:

- From
- Replay To
- Subject: è possibile customizzare il subject per notifiche relative all'host e ai servizi. Di default Check_MK propone una stringa parametrica che può essere modificata.

- Info da visualizzare nel body tra cui l'ip address dell'host o anche i dati di performance. In tal caso all'email vengono allegate le immagini dei grafici generati (Figura 32).

La definizione dell'oggetto di una email è un dettaglio fondamentale, in quanto è la prima cosa che un utente visualizza nella sua email, pertanto è importante che vi siano informazioni chiare come ad esempio il nome del server che ha generato la notifica, oppure il servizio monitorato o ancora il numero di server che hanno generato la notifica.

Figura 31-Sezione "Notification methods" di una regola di notifica.

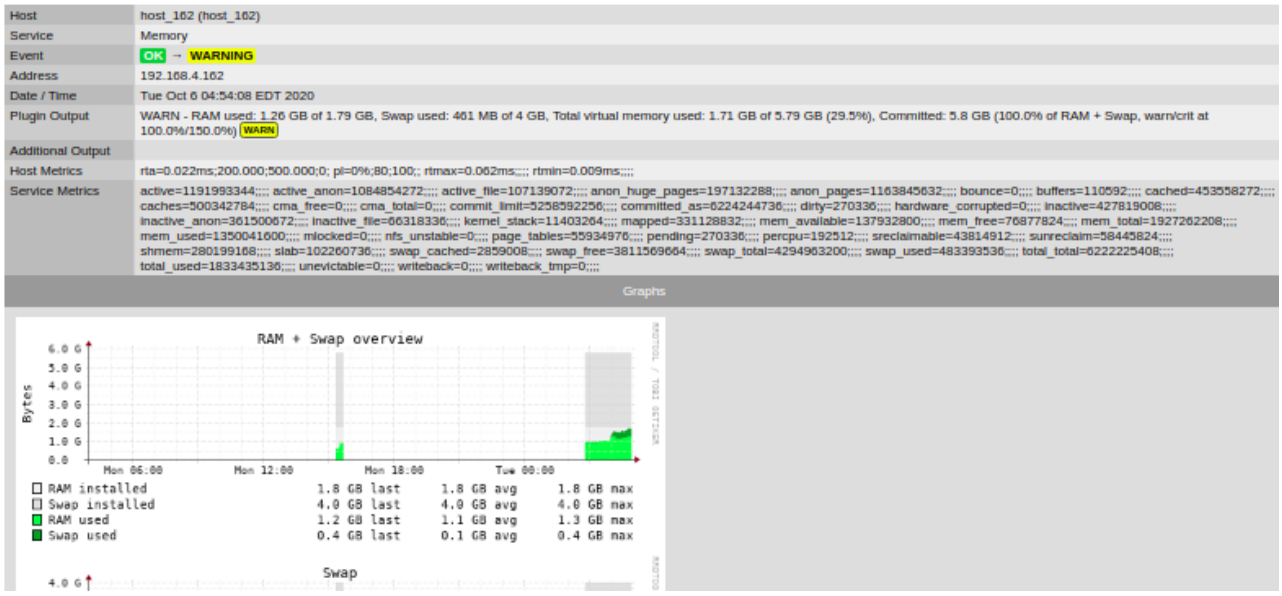

Figura 32- Un esempio di messaggio di notifica tramite email, che comprende anche i dati relativi alle performances in formato grafico.

Di particolare rilevanza è il flag "notification bulking" che permette il bulking delle notifiche, cioè l'invio massivo di un unico pacchetto di notifiche, qualora le stesse siano comprese in un intervallo, ad esempio 1 minuto. Le notifiche possono anche essere raggruppate separatamente per singolo host/servizio (vedi Figura

33). Il bulking delle email è utile per ridurre considerevolmente il numero di email che potrebbe essere abbastanza elevato in alcune situazioni.

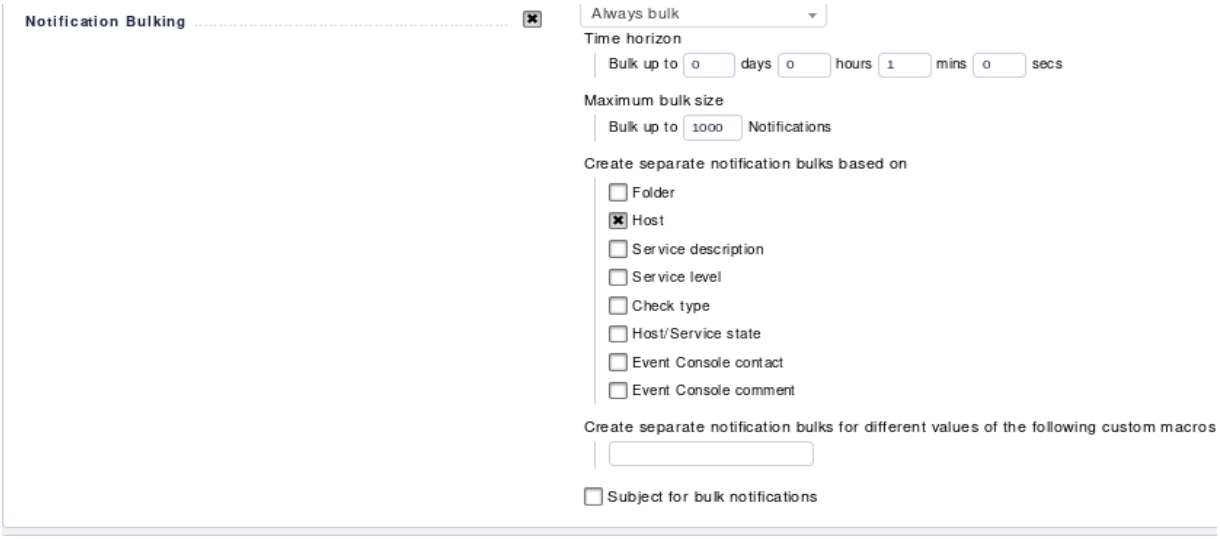

Figura 33-La sezione "Notification bulking" permette di utilizzare la modalità di invio massiva, al fine di ridurre o raggruppare notifiche in uno stesso messaggio.

Dalla sezione "contact selection" (Figura 34) è possibile selezionare a quali contatti inviare la notifica, ad esempio ad un gruppo di utenti.

È possibile notificare le email a tutti i contatti che hanno accesso ad un host o al servizio che si sta monitorando ("All contacts of the notified object"), oppure a tutti gli utenti; è possibile anche aggiungere utenti non presenti nei contatti ("the following explicit email addresses") oppure inviare l'email a singoli contatti ("The following users"). È possibile inoltre restringere l'invio a determinati gruppi usando la funzionalità "the members of certain contact group", definendo almeno il gruppo per cui applicare la regola.

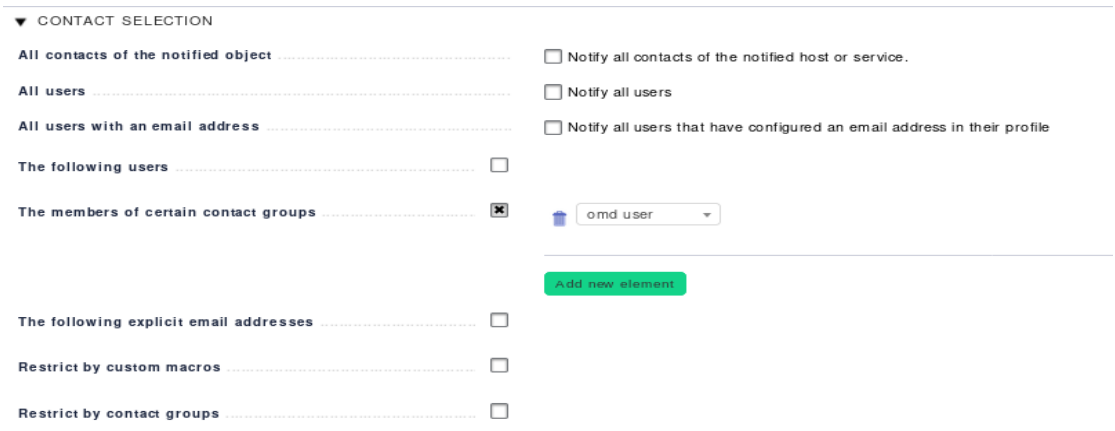

Figura 34-Sezione "contact selection" di una notifica, in cui è possibile indicare quali utenti debbano ricevere il messaggio.

Nella sezione successiva, "Conditions" (Figura 35 e 36), è possibile indicare le condizioni in seguito alle quali deve essere applicata la regola.

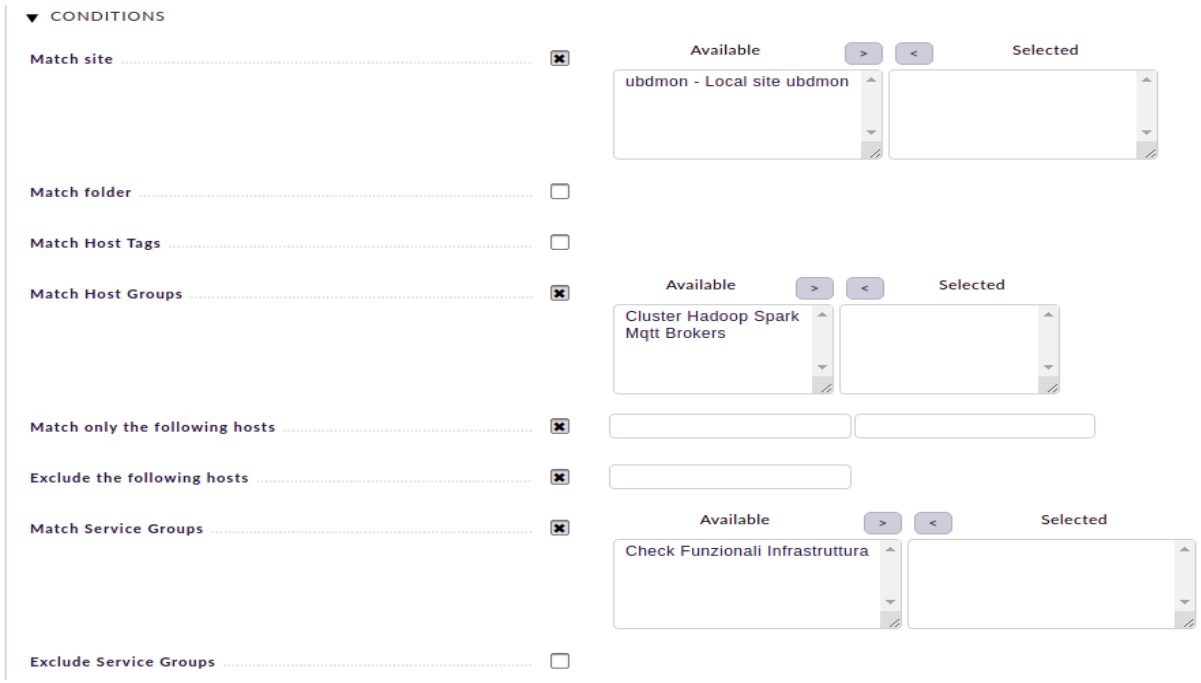

Figura 35-prima parte della sezione"Conditions" di una notifica. Qui è possibile definire gli hosts, oppure i gruppi di host, i servizi, oppure i gruppi di servizi, a cui applicare la regola di notifica.

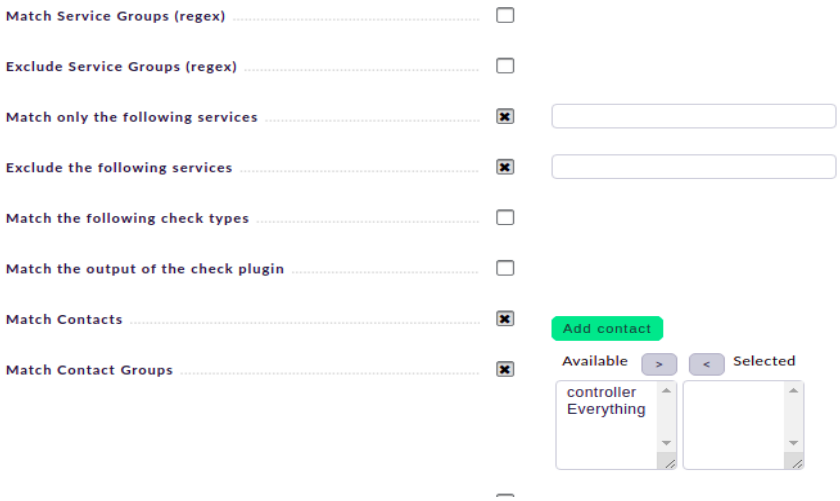

Figura 36-Seconda parte della sezione"Conditions" di una notifica. Qui è possibile definire anche i contatti o i gruppi di contatti che possono ricevere la notifica.

La definizione di una regola è una cosa delicata; infatti si deve evitare che vengano generate troppe notifiche, che farebbero perdere visibilità alle notifiche "vere", tuttavia non devono essere eccessivamente limitate perché altrimenti si rischia di perdere di vista eventuali problemi. Per definire come/quante notifiche configurare è necessario partire dalla configurazione delle notifiche per servizi/host più rilevanti, modificando poi le configurazioni in seguito ai rilevamenti.

Nelle condizioni è possibile specificare:

- Uno o più hosts da monitorare
- Uno o più servizi

- Gruppi di hosts eventualmente definiti (sono degli insiemi di hosts raggruppati secondo le necessità, e che ai fini del monitoraggio sono considerati come un unico host)
- Gruppi di servizi eventualmente definiti (insieme di servizi raggruppati)

Allo stesso modo è possibile escludere host o servizi dal monitoraggio ("exclude the following hosts" o "exclude Services Groups" o "exclude the following Services").

Per i server da noi monitorati, si è deciso di partire da alcuni servizi importanti, tra cui:

- Verifica se un host è attivo o no
- Verifica di servizi di sistema come CPU, RAM o disco
- Verifica di servizi specifici raggruppati ("gruppi di servizi" o "Service Groups")

Di seguito si riportano alcuni esempi di compilazione di questa sezione, che sono stati implementati da noi.

4.2.1 ESEMPIO 1: host attivo

Una informazione di interesse riguarda lo stato di attività di un host che, a causa della connessione o di problemi sistemistici (es firewall, black list, ecc.), può non essere più "visibile". Quando un host è in stato "DOWN", le notifiche relative ai suoi servizi non vengono inviate.

Dal momento che dobbiamo gestire un tipo di evento relativo all'host, il flag "Match host event type" nella sezione "conditions" ci aiuta a definire gli eventi da gestire.

Ad esempio, la Figura 37 mostra la configurazione della notifica per un evento relativo allo stato di un Host. Nell'esempio si è scelto di notificare all'utente quando un host subisce un evento del tipo selezionato nella lista; quando l'host passa da uno stato DOWN ad uno stato UP, vuol dire che prima l'host era irraggiungibile e successivamente è tornato on line. Selezionando any > DOWN, l'utente riceverà una notifica quando l'host passa da un qualsiasi stato(UP,UNREACHABLE) a DOWN.

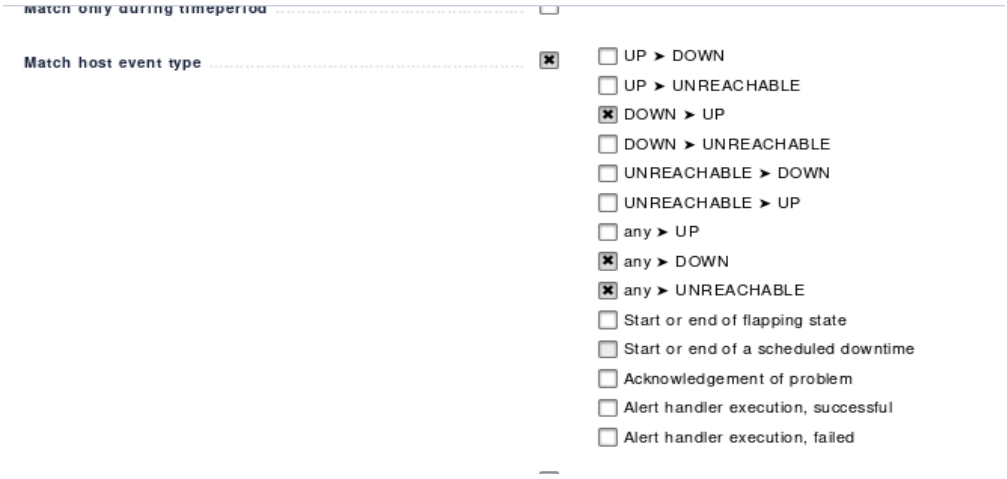

Figura 37- configurazione della notifica per un evento relativo allo stato di un Host.

Dal momento che questo controllo dovrebbe avvenire per tutti gli hosts, non è necessario selezionare altre informazioni. In caso si volessero escludere degli hosts dal controllo, è necessario selezionare il flag "exclude the following hosts" indicando gli hosts da eliminare. In caso invece si volesse applicare la regola solo ad alcuni hosts è necessario selezionare il flag "match only the following hosts" indicando gli hosts da controllare. Poiché OMD prevede la possibilità di raggruppare più server al fine di considerarli un unico host (group host), è possibile selezionare il flag "match hosts groups" indicando i gruppi di hosts configurati in precedenza da controllare. Effettuando il salvataggio, la notifica viene attivata. La Figura 38 mostra un

esempio di email html generato dalla configurazione indicata. In essa viene indicato l'host che ha subito l'evento, il tipo di evento(UP>DOWN, cioè è uscito dalla rete di monitoraggio per, ad esempio, problemi di connessione), l'indirizzo ip dell'host, la data e l'ora dell'evento, l'output generato dal check che ha scatenato la notifica(rta nan, lost 100% vuol dire che tutti i pacchetti scambiati con la macchina sono andati persi generando una situazione "CRITICAL").

Check_MK: host02_162 - UP -> DOWN

Host	host02_162 (host02_162)
Event	UP → DOWN
Address	192.168.4.162
Date / Time	Wed Oct 7 04:36:54 EDT 2020
Plugin Output	CRITICAL - 192.168.4.162: rta nan, lost 100%
Metrics	rta=0.000ms;200.000;500.000;0; pl=100%;80;100;; rtmax=0.000ms;;;; rtmin=0.000ms;;;;

Figura 38-Esempio di email HTML per l'esempio 1.

4.2.2 ESEMPIO 2: controllo di servizi specifici relativi ai sistemi (memoria, cpu, disco)

Check_mk prevede la possibilità di monitorare singoli servizi specifici per uno o più hosts; ciò avviene attraverso apposite funzioni che verificano eventuali cambi di stato dei servizi, e possono essere configurati in modo da rilevare specifiche situazioni più o meno critiche.

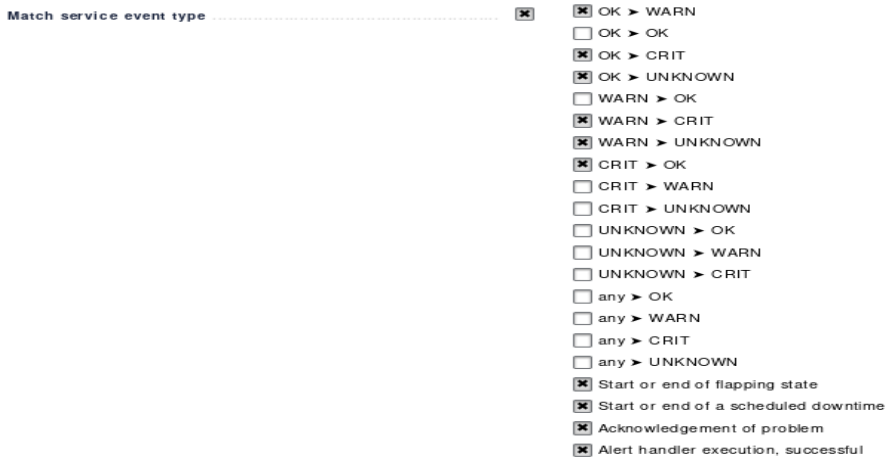

Figura 39-Lista dei cambi di stato per eventi che intervengono sui singoli servizi.

A tal fine è necessario selezionare il flag "Match service event type", definendo i cambi di stato che si desidera monitorare.
La Figura 39 mostra la lista dei cambi di stato con possibilità di selezionare solo quelli di interesse. Similarmente alla selezione degli eventi riguardanti gli hosts, in questa sezione è possibile selezionare gli eventi che intervengono sui singoli servizi, decidendo quali eventi siano da notificare all'utente. Ad esempio, se un check sullo stato del disco fisso controlla che le dimensioni di riempimento dell'hard disk debbano determinare una condiione di warning quando raggiungono il 60%, e una situazione di criticità quando raggiungono il 70%, nel primo caso il servizio subisce un cambiamento di stato passando da OK a WARNING, nel secondo caso passa da OK o WARNING a CRITICAL. In entrambi i casi viene mandata una notifica.
La regola così definita viene applicata a tutti i servizi e a tutti gli hosts come per la regola in esempio 1.

In caso si volessero escludere degli hosts dal controllo, è necessario selezionare il flag “exclude the following hosts” indicando gli hosts da eliminare.

In caso invece si volesse applicare la regola solo ad alcuni hosts è necessario selezionare il flag “match only the following hosts” indicando gli hosts da controllare.

In caso si volesse applicare ad uno o più gruppi di host definiti, è necessario selezionare il flag “match hosts groups” indicando i gruppi da controllare.

Per indicare in dettaglio quali servizi monitorare, deve essere selezionato il flag “match only the following services”, indicando i nomi dei servizi.

Nell’esempio in Figura 40 si è scelto di applicare la regola solo ai servizi Check_MK,CPU_utilization, Disk IO Summary e Memory. Quando il check introdotto su uno di questi servizi soddisfa la condizione indicata nella Figura 39, cioè si verifica un cambio di stato selezionato, viene inviata una notifica all’utente. Un qualsiasi cambio di stato su un servizio differente da quelli selezionati non genera alcuna notifica.

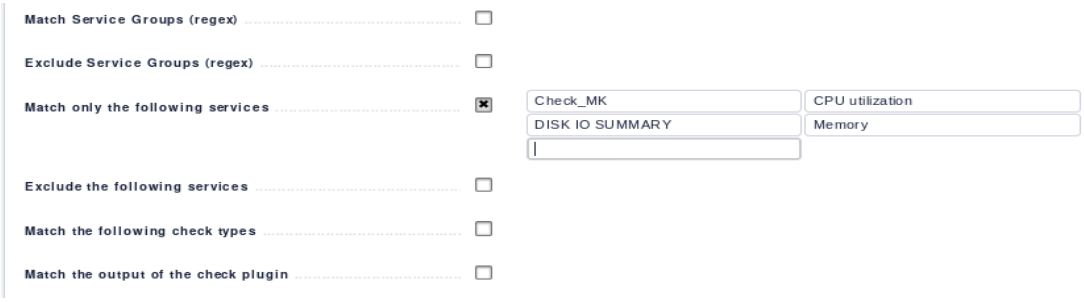

Figura 40-Applicazione della regola solo ai servizi Check_MK,CPU_utilization, Disk IO Summary e Memory.

Effettuando il salvataggio la notifica viene attivata. La notifica via email sarà simile alla seguente mostrata in Figura 41, in cui due eventi sono raggruppati in una notifica. Il primo evento mostra un cambio di stato per l’host01, per il servizio relativo al check sull’utilizzo della CPU che passa da OK a CRITICAL. Dopo circa 30 secondi il servizio ha nuovamente cambiato stato ritornando in stato OK, pertanto il sistema di notifica ha accorpato i due messaggi.

Nel messaggio sono indicati anche la data, l’ora e il messaggio (Plugin output) di output del check.

Check_MK: 2 notifications for host01.

Host	host01 (host01)
Service	CPU load
Event	OK → CRITICAL
Address	192.168.4.175
Date / Time	Wed Oct 7 03:21:19 EDT 2020
Plugin Output	dvaervse
Host Metrics	rta=0.011ms;200.000;500.000;0; pl=0%;80;100;; rtmax=0.039ms;;;; rtmin=0.004ms;;;;
Service Metrics	
Host	host01 (host01)
Service	CPU load
Event	CRITICAL → OK
Address	192.168.4.175
Date / Time	Wed Oct 7 03:21:55 EDT 2020
Plugin Output	OK - 15 min load: 0.58
Host Metrics	rta=0.011ms;200.000;500.000;0; pl=0%;80;100;; rtmax=0.039ms;;;; rtmin=0.004ms;;;;
Service Metrics	load1=1.17;5;10;0;1 load5=0.65;5;10;0;1 load15=0.58;5;10;0;1

Figura 41-esempio di email html generata per due eventi raggruppati in una notifica.

4.2.3 ESEMPIO 3: singoli eventi relativi ad un host specifico: WEB GUI

WEB GUI è un host definito in OMD che comprende i servizi relativi alle interfacce grafiche per i server che le prevedono, ad esempio il PELL.

Per gestire le notifiche relative ai singoli eventi di un host specifico, nel nostro caso WEB-GUI, si deve selezionare “Match service event type”, definendo i cambi di stato che si desidera monitorare.

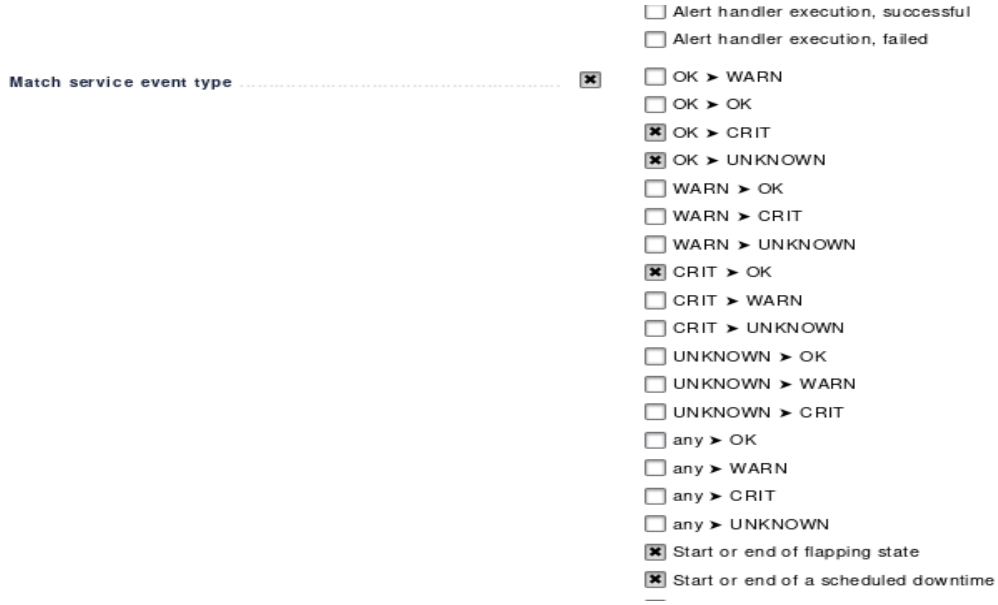

Figura 42-Esempio di configurazione degli stati di un evento.

Nell’esempio in Figura 42 è mostrata la configurazione degli stati di un evento. Si sta indicando di generare una notifica, quando perl’host WEB GUI, che comprende i servizi legati alle interfacce grafiche, uno dei servizi passa da uno stato OK a CRITICAL o UNKNOWN, e quando lo stesso ritorna ad uno stato di OK da uno stato CRITICAL.

Poiché si desidera monitorare dei servizi specifici di un host (WEB GUI), si deve selezionare “match only the following hosts” inserendo il nome dell'host “WebGui”. In Figura 43 si sceglie di monitorare l’host WEB GUI, e i servizi indicati nei campi “Match only the following services”. Indicando nessun servizio vengono generate notifiche per tutti i servizi presenti sull’host.

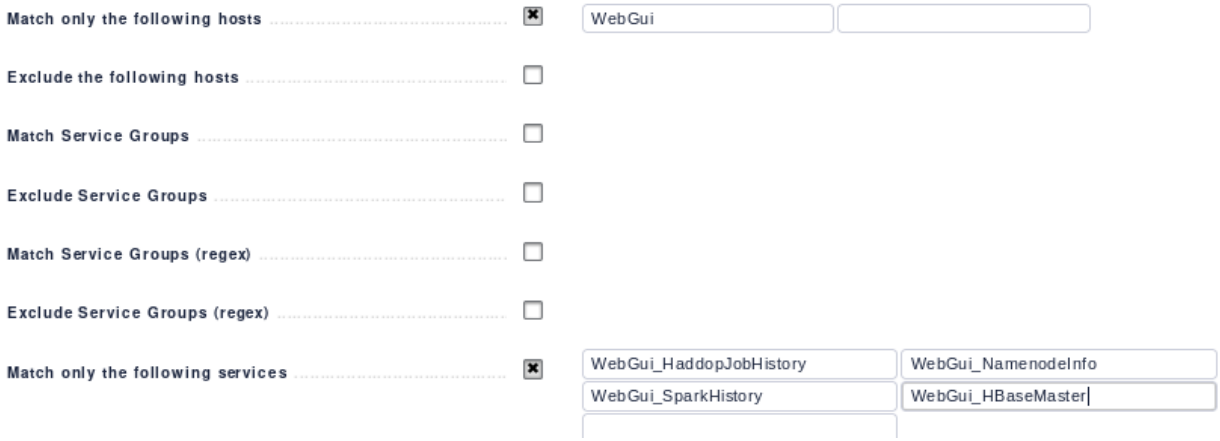

Figura 43-Definizione dei servizi singoli da monitorare.

Selezionando il flag “match only the following services”, possono essere indicati i servizi da monitorare, nel nostro caso:

- WebGui_HadoopJobHistory
- WebGui_NamenodeInfo
- WebGui_SparkHistory
- WebGui_HbaseMaster
- WebGui_PellInterno

Il salvataggio della regola determina la sua attivazione.

4.2.4 ESEMPIO 4: singoli eventi relativi ad un host specifico: Namenode

In questo esempio si vogliono generare notifiche in caso di cambio di stato per i servizi relativi ad HDFS e Yarn sull’host Namenode.

Come per l'esempio 3, si deve selezionare “match only the following hosts” inserendo il nome dell'host “Namenode”.

Selezionando il flag “match only the following services”, possono essere indicati i servizi da monitorare:

- Check_hdfs_put
- Check_Yarn

4.2.5 ESEMPIO 5: notifiche relative ai cambi di stato per i servizi dell’host “Servizi”

Come per l'esempio 3, si deve selezionare “match only the following hosts” inserendo il nome dell'host “Servizi”.

Selezionando il flag “match only the following services”, possono essere indicati i servizi da monitorare:

- LoadBalancer_Mqtt
- Pellbroker_Mqtt
- Pellbroker_Mqttssl

Gli esempi 3,4,5 sono stati descritti separatamente per chiarire la definizione di una regola di notifica in varie situazioni.

Su OMD è stato configurato un “Service Group” chiamato “Funzionali”, che come un “Host group”, raggruppa vari servizi, nel nostro caso:

- Check_hdfs_put
- Check_Yarn
- LoadBalancer_Mqtt
- Pellbroker_Mqtt
- Pellbroker_Mqttssl
- WebGui_HadoopJobHistory
- WebGui_NamenodeInfo
- WebGui_SparkHistory

- WebGui_HbaseMaster
- WebGui_PellInterno

  La figura 44 mostra il Service group “Funzionali”, che comprende più servizi, anche relativi a server differenti, raggruppati sotto lo stesso nome. In questo modo è possibile applicare una sola notifica per tutti i servizi.

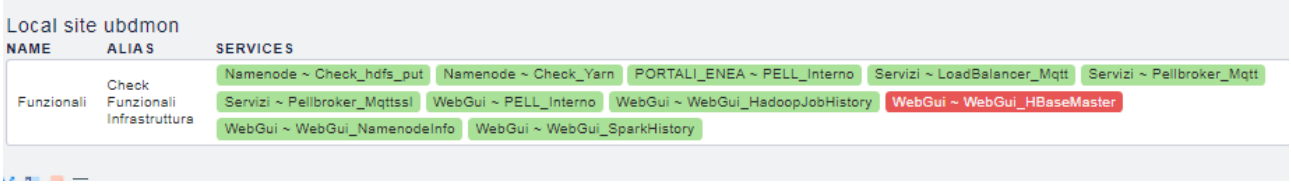

Figura 44-Il Service group “Funzionali”, che comprende più servizi, anche relativi a server differenti, raggruppati sotto lo stesso nome.

Questo permette la creazione di una notifica ad hoc, invece di 3 notifiche differenti.

Per identificare la condizione, basta selezionare “Match service groups” e il gruppo di servizi che si vuole monitorare. (Figura 45) In tal caso non è necessario selezionare alcun host, in quanto la notifica viene gestita direttamente sui servizi.

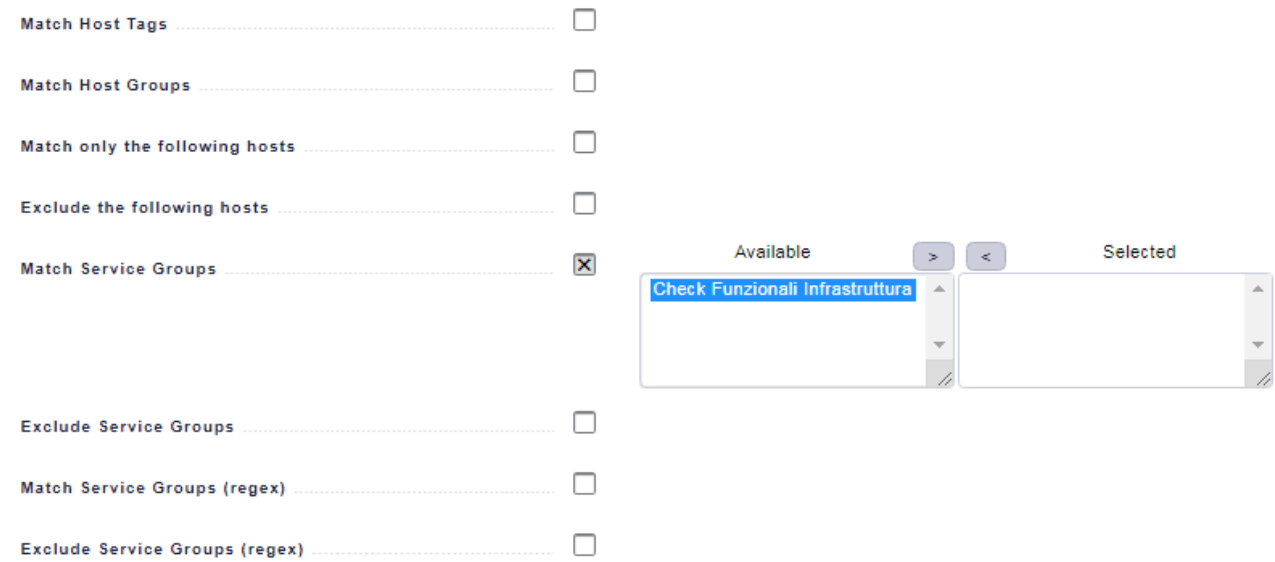

Figura 45-Viene selezionato il gruppo di servizi “Check funzionali infrastruttura, comprendente più servizi dello stesso tipo per cui può essere inviata la stessa notifica.

Per ulteriori informazioni sul sistema di gestione notifiche di Check_MK si può fare riferimento alla documentazione ufficiale al link seguente: CheckMK

5 Studio di algoritmi per l'analisi e il monitoraggio dei consumi di energia elettrica

Nell'ambito della collaborazione per lo sviluppo del calcolo dei KPI di tipo "dinamico", ci siamo occupati dell'analisi formale dei dati provenienti dagli smart meters al fine di immaginare quale potesse essere una strada percorribile per poterli trattare ed ottenerne ulteriori informazioni. Ad esempio, per un più efficiente monitoraggio dei consumi di energia elettrica, abbiamo cercato di analizzare la letteratura per individuare quali algoritmi, metodologie e strumenti potessero qualificare l'informazione contenuta nei dati grezzi provenienti dagli smart meters.
Il risparmio energetico è un punto centrale delle politiche governative di molti paesi per le implicazioni politiche ed economiche che ne derivano. Il tema pone molteplici sfide scientifiche e tecnologiche ed è al contempo frizzante sia dal punto di vista della produzione scientifica sia per ciò che concerne le applicazioni, i casi di studio e le implementazioni rintracciabili in tutto il mondo. Ciò si spiega con il fatto che molte delle tecnologie impiegate sono economicamente accessibili e, in tempi medi, diventano convenienti. E' il caso delle lampade a LED che, a fronte di un costo iniziale più elevato rispetto alle lampade costruite con tecnologie classiche, offrono efficienza molto più elevata e quindi, in tempi ragionevoli, permettono il recupero dell'investimento potendo contare su costi legati ai consumi energetici inferiori. Il risparmio energetico può essere ottenuto attraverso step successivi come appunto l'impiego di tecnologie innovative (caso delle lampade al LED) che aumentano l'efficienza dell'impianto, o con l'impiego di tecnologie di controllo dell'accensione/spegnimento degli impianti. In tutti i casi, per avere contezza della situazione ante e post opera e quindi per giustificare l'investimento ed al contempo verificarne i benefici in termini di minori consumi, sostenibilità ambientale e risparmi economici, non si può prescindere da un monitoraggio dei consumi. La sedimentazione delle misure sui consumi può poi permettere ragionamenti sulla predizione del comportamento del sistema (o dei sistemi) sotto indagine, in modo da procedere ad una progettazione/programmazione di interventi migliorativi.
La predizione avviene attraverso lo studio e l'analisi di modelli numerici a ciò preposti come gli algoritmi predittivi o attraverso l'analisi statistica dei dati di consumo.
In previsione di sostenere il lavoro ENEA al fine di potenziarne le possibilità in termini di efficienza, l'oggetto della nostra indagine è stato proprio l'individuazione di algoritmi e metodologie usate per le predizioni dei consumi energetici ottenute attraverso l'analisi della letteratura. Essendo la previsione dei dati e ancor di più l'analisi dei dati, temi largamente dibattuti nella comunità scientifica, l'approccio usato ha previsto di ripercorrere la letteratura per i metodi di analisi e previsione ampiamenti usati in molti settori e riportati nei paragrafi dal 5.1 al 5.9 supportandola sia con richiami bibliografici generali sia con lavori che abbiamo ritenuto più significativi per rigore metodologico o perché mettono in luce tendenze innovative particolarmente stimolanti. Nel paragrafi vengono richiamati alcuni articoli, una selezione limitata della letteratura disponibile in merito, che mostrano anche come i metodi descritti da un punto di vista analitico possano trovare applicazione pratica nella gestione dei dati energetici e nelle previsioni di consumi elettrici. L'attività impostata quest'anno sarà poi oggetto di lavoro ed implementazione durante il terzo anno.
In questo capitolo abbiamo descritto dapprima alcune tra le metodologie di analisi dei dati (5.1), tra cui i metodi statistici e matematici:

- L'analisi predittiva(5.2)
- I modelli classici (5.3)
- I modelli di regressione (5.4)
- I modelli ARIMA (AutoRegressive Integrated Moving Average) (5.5)

E quelli basati su intelligenza artificiale e Machine learning:

- Le reti neurali(5.8),

che possono comprendere anche gli algoritmi di clustering come

- il K-means (5.6)
- SVM(support vectror machine) (5.7),

attualmente tra i metodi più utilizzati.

La Figura 46 mostra le tecniche di analisi implementabili, da una parte i metodi basati esclusivamente su approcci matematici e statistici (paragrafi da 5.2 a 5.5) dall'altra i metodi basati sull'intelligenza artificiale (paragrafo 5.6) ed in fine i metodi di clustering che possono supportare i precedenti (paragrafi 5.7 e 5.8).

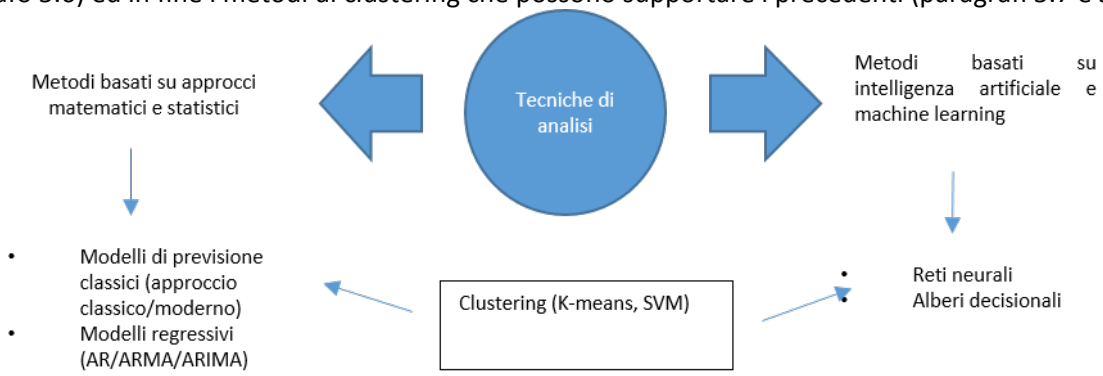

Figura 46–Tecniche di analisi dei dati suddivise in metodi matematici/statistici e metodi basati su intelligenza artificiale.

Successivamente abbiamo descritto brevemente i dati in nostro possesso (paragrafo 6.9), evidenziando il fatto che, per quantità e struttura, essi ricadono nel mondo dei "Big data", cioè di grandi quantità di dati per i quali è necessario avvalersi di strumenti dedicati all'archiviazione e all'elaborazione. Lo studio di questi dati attraverso le tecniche e metodologie tipiche dei Big Data rappresenta una novità importante e potrebbe essere foriera di una linea di ricerca innovativa.
La letteratura è molto ricca di pubblicazioni e ricerche scientifiche che indicano quale potrebbe essere il percorso da seguire per il nostro caso specifico. Da queste ricerche è emerso come il tema dei consumi energetici sia stato affrontato da molti punti di vista come, ad esempio, l'analisi di particolari scenari, la previsione di consumi o di domanda di energia da parte di alcuni settori come quelli oggetto dei nostri studi, cioè l'illuminazione pubblica e il consumo di edifici scolastici.
L'analisi della letteratura è dunque un'attività preliminare fondamentale perché pone le basi per l'analisi dei dati energetici in nostro possesso ed attraverso lo studio dei vari tipi di modelli e delle modalità di utilizzo di ogni metodologia, è possibile valutare la loro implementazione con riferimento al nostro particolare scenario. Infatti, non esiste un'unica modalità per l'analisi dei dati, ma ogni tecnica o metodo (che spesso sono utilizzati anche in combinazione) può mostrare una particolare caratteristica del fenomeno, favorendone l'analisi e la previsione.
Nello studio degli algoritmi di analisi dei dati abbiamo tenuto particolarmente in considerazione lo scenario applicativo che prevede un set di dati che, per la sua dimensione presenta aspetti non banali. Infatti la potenziale quantità dei dati a disposizione rende questo data set enorme facendolo ricadere in un ambito particolare di grande attualità dal punto di vista tecnico scientifico e che prende il nome di "Big Data".
La gestione dei "Big Data" è una attività relativamente recente e lo è ancor di più nel framework del risparmio energetico impostato da ENEA; l'analisi da noi condotta ha fatto emergere che a dispetto del carattere innovativo di questa impostazione, essa può avvalersi pienamente di metodi di analisi dei dati di tipo classico creando un connubio di cui inizialmente non si aveva certezza. L'evoluzione rispetto al passato è la tecnologia attuale insita nei concetti legati ai Big Data che ci viene in aiuto nella gestione dell'enorme quantità di dati di cui si potrebbe disporre e che, se fornita tout court come input ai metodi di analisi classica, ne impedirebbero un uso efficiente e quindi produttivo. Infatti, l'analisi dei dati dipende dalla loro numerosità, aspetto che incide fortemente da un punto di vista computazionale e che attualmente viene affrontato in vari modi, ad esempio gestendo i dati su cloud, oppure attraverso piattaforme dedicate che facilitano la storicizzazione, il recupero e l'elaborazione di dati, il cui ordine di grandezza, a regime, è dell'ordine dei PetaByte. E' quindi immaginabile che le metodologie di analisi dei dati classica possano essere espanse potendo attingere ai paradigmi dei Big Data.

Sulla base di questa ricerca, forti delle considerazioni fatte sullo scenario e sul data set, nel paragrafo 5.10 azzardiamo una proposta operativa che, sebbene considerabile embrionale, rappresenta la traccia di quello che dovremmo implementare nella terza annualità.
Verranno poi elencati alcuni strumenti (paragrafo 5.11) per la gestione e l'elaborazione di dati ed, infine, nelle conclusioni, riassumeremo i risultati delle nostre analisi, i metodi e gli strumenti che sembrano più adeguati al nostro scenario.

5.1 Tecniche di analisi

In questa sottosezione, il nostro contributo è stato quello di analizzare la letteratura cercando di presentare in modo razionale gli strumenti tipicamente usati per effettuare analisi dei dati relativi a consumi energetici cercando anche di evidenziarne tendenze ed innovazioni recenti. Infatti, per effettuare questi tipi di analisi esistono vari strumenti, ciascuno maggiormente efficace se contestualizzato nel proprio ambito; ne riportiamo qui alcune categorie: [4]

- Modelli di previsione classici: utilizzano alcuni indicatori economici e coefficienti specifici e sono tesi alla previsione del carico elettrico, cioè alla previsione del fabbisogno futuro di energia elettrica.
- Modelli econometrici, come ad esempio quelli che vengono utilizzati attualmente per prevedere il PUN (prezzo unico dell'energia elettrica) o metodi di regressione o modelli di serie temporali. I più usati in ambito elettrico sono proprio i metodi di regressione, che mettono in relazione la domanda di energia e altri fattori come, ad esempio, il tempo o il numero di abitanti.
  In questa tipologia di modelli spiccano in particolare:
  - I modelli basati sulle serie storiche.
    Questi sono modelli utilizzati molto spesso per la loro semplicità; infatti considerano solo i dati di input storici, che non sono messi in relazione con alcun fattore, e ne valutano l'andamento. Ovviamente la loro limitazione è proprio nella semplicità che esclude il loro utilizzo in sistemi complessi anche multivariabile.
  - I metodi basati sulla regressione lineare multipla.
    Questi modelli, anch'essi molto utilizzati, mettono in relazione i dati di input con fattori esterni, ma presentano dei limiti legati ai DataSets di input; infatti più sono affidabili e completi e, tipicamente, migliore sarà il risultato di previsione.
- Modelli basati su tecniche di intelligenza artificiale (ad esempio l'uso di reti neurali o algoritmi di machine learning).
  Questi modelli permettono di effettuare un salto di qualità nello studio previsionale, infatti, attraverso questi si possono affrontare modelli molto più complessi anche non lineari. Tuttavia, non richiedono grandi formulazioni matematiche o correlazioni particolari tra i dati. Ad esempio, le reti neurali, una volta configurate nel modo corretto, prendono in input i dati grezzi e li elaborano acquisendo informazioni e apprendendo in automatico.

Ciascuna modalità descritta ha delle particolari caratteristiche che la rende maggiormente adatta ad uno scenario piuttosto che ad un altro.
Una particolarità dello scenario che andiamo a studiare, è sicuramente la potenziale elevata numerosità dei dati che rende ancor più importante la definizione delle modalità e degli strumenti da utilizzare.
Da alcuni anni a questa parte nel mondo informatico e dell'analisi dei dati è emerso un particolare settore che va ad analizzare non solo il valore informativo del dato, ma anche le problematiche tecnico-scientifiche legate a grandi quantità di dati. Questo ambito è comunemente e sinteticamente chiamato nel lessico tecnico "Big Data".
Il campo dell'analisi dei big data, e in particolare dei big data relativi al mondo dell'energia, è di grande e vivo interesse da parte di molte organizzazioni, come università o enti del settore energetico (ad es. fornitori di energia elettrica).
Questo, vista la grande quantità e varietà di metodologie proposte, spesso usate anche in combinazione, rende la scelta degli algoritmi da utilizzare difficile, affermazione ancor più vera in considerazione della forte

variabilità dei dati che è importante al fine di stabilire sia la corretta metodologia di analisi, sia la definizione della la struttura del dataset che si desidera utilizzare.
Nel caso di dati temporali, sicuramente è possibile partire da **metodi di tipo quantitativo**, cioè quei metodi basati su tecniche statistiche che necessitano:

- Di una sufficiente quantità di dati "storici" che permettano di valutare l'andamento di un determinato fenomeno.
- Della supposizione che le caratteristiche dell'evoluzione del fenomeno si mantengano anche nel futuro.

Essi si dividono in metodi di analisi di serie storiche e metodi esplicativi.

- Una serie storica è una sequenza di dati che sono ordinati rispetto al tempo (ad esempio il consumo mensile o giornaliero di energia, il consumo di potenza, ma anche il costo annuo di manutenzione degli impianti, etc.).
  Le serie storiche considerano come unica variabile il tempo; l'evoluzione del fenomeno è legata all'evoluzione temporale, ed in base all'evoluzione passata si cerca di estrapolare un percorso per ottenere una previsione.
  Lo scopo dell'analisi delle serie storiche consiste nello studio dell'evoluzione passata del fenomeno rispetto al tempo, cercando di ottenere una previsione considerando che l'evoluzione nel futuro presenti la stessa modalità di comportamento.
  Generalmente questo tipo di analisi è utile quando si conosce poco il fenomeno e quando interessa conoscere quello che avverrà ma non la modalità.
- I metodi esplicativi utilizzano i metodi di regressione e mettono in relazione la variabile da prevedere con altre variabili "esplicative" che esplicitano l'andamento del fenomeno da prevedere. Se le variabili esplicative sono manipolabili dal decisore, questi metodi sono molto efficienti nella previsione del fenomeno.

5.2 Analisi predittiva

L'analisi predittiva si basa su alcuni passi fondamentali, in quanto si deve cercare di ricavare il modello migliore che possa fornire una previsione quanto più possibile accurata. Pertanto, per prima cosa è necessario definire la problematica che si va ad analizzare e poi è necessario comprendere cosa si deve prevedere e come tali previsioni verranno utilizzate. [5-7]
È necessario poi avere visibilità completa delle informazioni relative ai dati di input. Infatti spesso non basta conoscere il dato grezzo, ma anche eventuali fenomeni che lo condizionano. Un esempio è il consumo di energia che può variare a seconda del periodo dell'anno o delle condizioni climatiche oppure può essere legato al contesto in cui i dati sono raccolti, come nel caso di consumi legati all'illuminazione pubblica o ad edifici scolastici o ancora ad abitazioni private.
Questo modo di procedere è importante in fase di prima analisi dei dati. Infatti, al fine di prevedere un modello corretto, devono essere analizzati i dati in possesso attraverso metodi grafici, come diagrammi di scattering o time plot, oppure vari tipi di scomposizioni di dati che permettono di recuperare o eliminare informazioni di anomalie, irregolarità, situazioni particolari oppure cicliche. La Figura 47 mostra un esempio di time plot, che rappresenta i dati rispetto all'evoluzione temporale.
Attraverso un time plot è possibile rilevare particolarità nell'andamento dei dati, evidenziando dei trend di evoluzione temporale ed eventualmente, attraverso l'applicazione di metodi matematici quali il calcolo della media, è possibile determinare l'andamento della serie. In particolare:

- Quando una serie presenta dei valori che oscillano attorno alla sua media, si parla di **serie stazionaria** e di **trend orizzontale** della serie.
- Quando la serie presenta delle periodicità, ad esempio in caso di stagionalità, mesi, anni ecc., si parla di **serie periodiche** e di **trend stagionale**.
- Quando la serie oscilla in periodi che non sono fissi come nel caso di trend stagionale, si parla di **serie cicliche** o **trend ciclico.**

- Infine, quando una serie presenta un andamento crescente o decrescente nel lungo periodo, si parla di **trend residuale**.

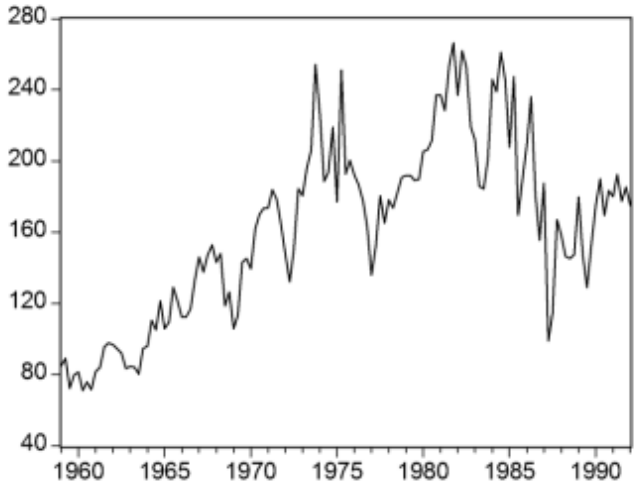

Figura 47- esempio di time plot.

Questi trend, normalmente, non sono esclusivi l'uno dell'altro, infatti è possibile che in una stessa serie si possano distinguere più trend, e un buon modello è in grado di riconoscerli.
Oltre all'aiuto che arriva dalla visibilità che può fornire un grafico, è necessario effettuare una analisi descrittiva dei dati, cioè lo studio degli indici statistici come:

- La media, cioè la somma di tutti i valori delle variabili della popolazione di dati (y_i) diviso il numero di unità della popolazione di dati:

$$\bar{y} = \frac{1}{n}\sum_{1}^{n} y_i$$

- La moda che indica la modalità caratterizzata dalla massima frequenza, cioè il valore che compare più frequentemente.
- La mediana che è un indice di posizione che restituisce il valore mediano di una distribuzione. Generalmente è poco influenzata da dati particolari.
- Il campo di variazione che è la differenza tra il dato (y_i) più grande e quello più piccolo della distribuzione:

$$R = Max(y_i) - Min(y_i)$$

- La varianza che fornisce una misura della variabilità dei dati essendo definita come la media aritmetica dei quadrati delle differenze tra ogni valore y_i e un valore medio preso come riferimento:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

- La deviazione standard o scarto quadratico medio, esprime la dispersione dei dati intorno alla media (la formula è data dalla radice quadrata della formula precedente):

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

- La covarianza è un indice che permette di verificare una relazione lineare tra due variabili, cioè il valore atteso dei prodotti delle distanze di due variabili y dalla media:

$$c_k = \frac{1}{n-1}\sum_{i=1}^{n-k}(y_i - \bar{y})(y_{i+k} - \bar{y})$$

- La correlazione misura il grado di associazione tra due variabili, ovvero la variazione di una variabile rispetto all'altra. È definita come rapporto tra la covarianza delle due variabili e il prodotto delle loro deviazioni standard

$$r_k = \frac{\frac{1}{n-1}\sum_{i=1}^{n-k}(y_i - \bar{y})(y_{i+k} - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Un altro importante passo nell'analisi dei dati, è la possibilità di modificare i dati originari in modo da renderli meglio interpretabili. Questo avviene tramite trasformazioni dei dati, che possono essere:

- Matematiche
- Modifiche al fine di eliminare oscillazioni o valori che possono influenzare gli indici ma che possono essere trascurati perché poco indicativi.

Ad esempio I **metodi di scomposizione,** che possono essere **additivi**, **moltiplicativi**, o **misti**, permettono di separare le componenti casuali dall'andamento caratteristico di una serie storica favorendo una migliore accuratezza della previsione. [8]
Ciò avviene considerando la serie come costituita da tre parti:

- Il ciclo-trend che identifica sia l'andamento di fondo della serie in un arco temporale lungo (trend), sia le fluttuazioni (ciclo).
- Stagionalità, che identifica tutte le fluttuazioni che avvengono con una certa periodicità e dovute ad alcuni fattori specifici (ad esempio la temperatura, le stagioni, ecc.).
- Componente residuale, che identifica le variazioni casuali e non prevedibili.

È possibile dunque definire una funzione di scomposizione **Z**, dove **T** è la funzione relativa al trend, **C** è la funzione relativa al ciclo, **S** è la funzione che descrive la stagionalità, ed **A** è la funzione che rappresenta la componente residuale:

$$Z = T+C+S+A$$

Essa rappresenta un modello **additivo**, e presenta le seguenti caratteristiche:

- Tutte le variabili sono indipendenti.
- Tutte le variabili hanno la stessa unità di misura
- Il suo errore può assumere valori positivi, negativi o nulli (neutralità, cioè non influenza la serie).

Questo modello è appropriato quando l'ampiezza dell'oscillazione stagionale non varia con il livello della serie.

A differenza del modello additivo, il modello **moltiplicativo** presenta:

- Tutte le componenti dipendenti,
- T, C e Z hanno la stessa unità di misura, mentre S e A sono adimensionali
- Il suo errore può assumere solo valori maggiori o uguali a zero, essendo 1 la neutralità.

Trova applicabilità quando la fluttuazione stagionale varia proporzionalmente con il livello della serie:

$$Z = T \times C \times S \times A$$

In ultimo, il modello **misto** comprende invece il prodotto della componente trend con stagionalità, sommato al prodotto tra la componente ciclo e alla componente residuale:

$$Z = T \times S + C \times A$$

La scomposizione è una metodica utile per una prima analisi. Se effettuata in modo corretto, resta un valido strumento per comprendere le caratteristiche evolutive passate della serie storica.
La prima analisi dovrebbe fornire l'indicazione di specifici modelli ottimali per la previsione. Infatti, i modelli sono tantissimi, ed è necessario scegliere quello che centra meglio la problematica da affrontare.

Una volta definito il modello, si può usare per la previsione. La previsione deve essere poi valutata in merito alla sua bontà, infatti è necessario utilizzare degli indicatori che possano permettere di capire se il modello è corretto, e deve essere effettuato un riscontro con i dati reali per avere visibilità dell'effettiva bontà del lavoro, che in gergo tecnico prende il nome di accuratezza della previsione.
Quindi, con il termine accuratezza, indichiamo la capacità del modello di riprodurre i dati sui quali è stato stimato, e che quindi ne misura la validità.
Il termine accuratezza di una previsione indica quanto il valore previsto di una quantità si avvicina al suo valore reale, pertanto ci fornisce una stima quantitativa della qualità attesa da una previsione, e il suo valore ci aiuta a valutare la bontà del modello che stiamo utilizzando. Essa rappresenta la concordanza tra la media aritmetica dei valori ottenuti e il valore reale. Nel caso di una previsione il valore reale può essere ottenuto solo a posteriori, tuttavia ci sono alcuni metodi per valutare l'accuratezza.
Ad esempio:

1. Si dividono i dati del modello in due parti. La serie è indicata come:

$$y_1, y_2, y_3, \dots, y_n$$

2. Si utilizzano i primi M dati per la stima

$$y_1, y_2, y_3, \dots, y_m$$

   Dove m < n

3. Si utilizzano i restanti dati per la verifica dell'accuratezza

$$y_{m+1}, y_{m+2}, y_{m+3}, \dots, y_n$$

Le stime ottenute con i primi m campioni sono:

$$\hat{y}_1, \hat{y}_2, \hat{y}_3, \dots, \hat{y}_m$$

Le previsione ottenute sono:

$$P_{m+1}, P_{m+2}, P_{m+3}, \dots, P_n$$

L'errore di stima è definito quindi come la differenza tra il valore del dato e il suo valore stimato:

$$e_t = y_t - \hat{y}_t$$

con t = 1, …, m

Mentre l'errore di previsione è definito come la differenza tra il valore stimato e il valore della previsione

$$f_t = \hat{y}_t - P_t$$

con t = m+1, …, n
Gli indicatori per misurare l'accuratezza sono molti, ma quelli più utilizzati sono i seguenti:

- Errore medio, o mean error: è la media aritmetica degli errori, e può assumere anche un valore negativo. In caso di valore negativo, il fenomeno descritto dal modello risulta sopravvalutato, in caso di valore positivo risulta invece sottovalutato

e

$$ME = \frac{1}{m} \cdot \sum_{i=1}^{m} e_i$$

- Errore quadratico medio o mean square error: è la media aritmetica dei quadrati degli errori:

$$MSE = \frac{1}{m} \cdot \sum_{i=1}^{m} {e_i}^2$$

- Errore medio assoluto o mean absolute error: è la media aritmetica degli errori presi in valore assoluto:

$$MAE = \frac{1}{m} \cdot \sum_{i=1}^{m} |e_i|$$

- Errore medio assoluto percentuale o mean absolute percentage error: è la media aritmetica degli errori relativi, presi in valore assoluto e moltiplicati per 100.

$$MAPE = \frac{1}{m} \cdot \sum_{i=1}^{m} \frac{|e_i|}{y_i} \cdot 100$$

Talvolta è utile e necessario affiancare l'analisi di questi indici ad altre metodiche (analisi grafiche degli errori, esperienza degli operatori, ecc.) allo scopo di dare una migliore valutazione dell'accuratezza della stima o della previsione [9].

5.3 Modelli di previsione classici

L'approccio classico alla previsione si avvale dei modelli stocastici che tengono in considerazione le variazioni causali e non causali delle variabili di input, e restituiscono una previsione "probabilistica". Un modello stocastico considera la variabilità dei dati di input.
Il modello stocastico si differenzia dal modello deterministico in quanto quest'ultimo considera fisse le variabili di input, non tenendo conto dell'incertezza a loro associabile.
Un modello stocastico si basa sull'assunzione che una serie storica sia costituita da due parti:

- Una sequenza deterministica (considerata in un approccio tradizionale)
- Una sequenza di eventi casuali che rispondono ad una legge di probabilità (considerata in un approccio moderno).

Essenzialmente, dunque, questo modello divide in due parti la funzione che rappresenta i dati storici; la sequenza deterministica corrisponde ad una funzione, di cui la variabile è il tempo, che è l'oggetto effettivo dell'analisi dei dati, mentre la sequenza casuale viene studiata separatamente, essendo considerata associata ad eventi particolari oppure errori.
Nell'approccio tradizionale la sequenza di eventi casuali è considerata generata da una successione di variabili casuali indipendenti che possiedono varianza costante e media nulla, quindi trascurabile.
In un approccio più moderno, tale componente viene invece considerata secondo teorie probabilistiche, cercando una correlazione statistica tra le grandezze per cercare di determinare un rapporto di causa/effetto.

5.4 Modelli di regressione

I modelli di regressione cercano di determinare delle relazioni tra tutte le variabili di interesse. A differenza dei metodi classici che separano le due componenti deterministica e probabilistica cercando di prevedere l'andamento futuro sulla base dell'andamento passato, i modelli di regressione cercano delle relazioni tra le variabili e cercano di identificarne il significato.
In questo modo l'andamento temporale del fenomeno viene integrato da altri fattori; ogni fattore considerato genera un risultato differente che necessita di essere interpretato nel modo corretto. Questo permette di comprendere il fenomeno in modo più accurato, considerando molti degli eventi che lo costituiscono, ed eventualmente valutando come le combinazioni di questi eventi influiscono sulla previsione.
Quest'ultima caratteristica permette di prevedere, ed in qualche modo anche di influenzare, le decisioni attuali, stimando le modalità di interazione tra i dati di input e quelli di uscita, ottenendo sicuramente delle previsioni migliori. [5-7, 10,11]
Un fattore fondamentale che influenza questi metodi è sicuramente la scelta delle variabili di input e di output; la variabilità di questa scelta, determina previsioni differenti.
La regressione lineare è una tecnica di modellazione statistica, che viene utilizzata per descrivere una variabile di risposta (y), in funzione di una o più variabili indipendenti dette predittori (x_i). Viene quindi creato un modello lineare a partire da un modello complesso.
Questo modello lineare assume la seguente forma matematica:

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i \cdot x_i + \varepsilon_i$$

in cui i coefficienti β_0, β_i rappresentano le stime per i parametri lineari da calcolare, mentre ε_i rappresenta i termini di errore.
La regressione lineare semplice utilizza un unico predittore x_1:

$$y = \beta_0 + \beta_1 \cdot x_1 + \varepsilon_i$$

che si traduce nell'immagine in Figura 48 in cui viene mostrato come la retta disegnata approssimi l'andamento del fenomeno.

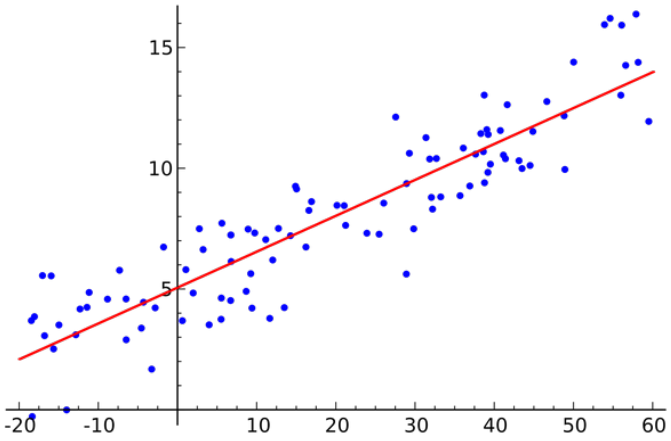

Figura 48-esempio di grafico relativo ad una regressione lineare semplice. I punti blu indicano i valori intorno alla retta che approssima linearmente la funzione.

La regressione lineare multipla utilizza più predittori:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \varepsilon_i$$

Con il risultato mostrato in Figura 49 in cui è visibile un piano di regressione anziché una retta.

Per la stima dei parametri $\beta_0, \beta_1, \beta_2$ si utilizza il metodo dei minimi quadrati che minimizzano l'errore e permettono di trovare la curva che minimizza la somma dei quadrati delle distanze tra i dati osservati e quelli della curva che rappresenta la funzione.

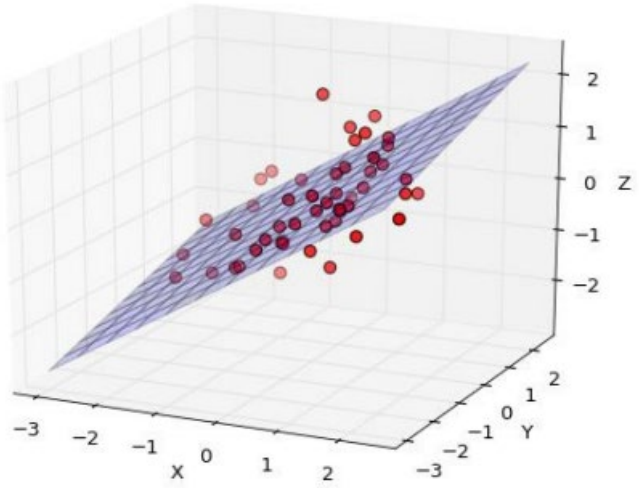

Figura 49—esempio di grafico relativo ad una regressione lineare multipla. I punti rossi sono i valori intorno al piano che approssima la funzione.

Il termine di errore ε descrive una variazione imprevedibile o casuale della variabile dipendente e deve essere anch'esso stimato.

La regressione lineare multivariata, infine, prevede modelli per variabili multiple di risposta, pertanto i modelli vengono indicati nel seguente modo:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \beta_{00} & \beta_{01} \\ \beta_{10} & \beta_{11} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ x_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

5.5 Arima

Il modello ARIMA (AutoRegressive Integrated Moving Average) è tra i modelli più comuni per l'analisi dati e realizzazione di forecasting.[5,12-13]

Esso è utile per l'analisi di dati di serie storiche e comprende tre modelli:

- AR o auto regressive che relaziona i dati passati con i dati successivi
- MA o moving average che imposta l'errore del modello come una combinazione lineare dei valori di errore osservati in punti temporali precedenti al passato.
- Una parte integrata I che gestisce la differenza di una osservazione da una osservazione nella fase temporale precedente, e ci serve per rendere stazionaria la serie temporale.

Il modello auto regressivo AR si basa sul presupposto che un fenomeno presente è proporzionale al fenomeno passato a cui si aggiunge un disturbo.

Può essere descritto come:

$$y_t = \beta_0 + \beta_1 \cdot y_{t-1} + \varepsilon_t$$

Che rappresenta un modello AR di livello 1, dove y_t è il valore della serie storica al tempo t, y_{t-1} è il valore della serie storica al tempo t-1, i coefficienti β sono costanti che descrivono la funzione di primo grado, mentre ε_t rappresenta un "white noise", cioè una gaussiana. Una proprietà di questo modello, è legata

proprio agli scarti, che non sono indipendenti tra loro e permettono l'uso di metodi di massima verosimiglianza.
In un sistema di questo tipo, devono essere stimati i parametri β e questo prevede che i processi siano stazionari, ovvero la serie storica deve avere la media e la varianza costanti (Figura 50).

Figura 50-Esempio grafico di una serie stazionaria.

Quando il processo non è stazionario ma "esplosivo" (Figura 51), la stima dei parametri non è più affidabile, e richiede un ulteriore step che prende il nome di integrazione e che risulta in una trasformazione dei dati.

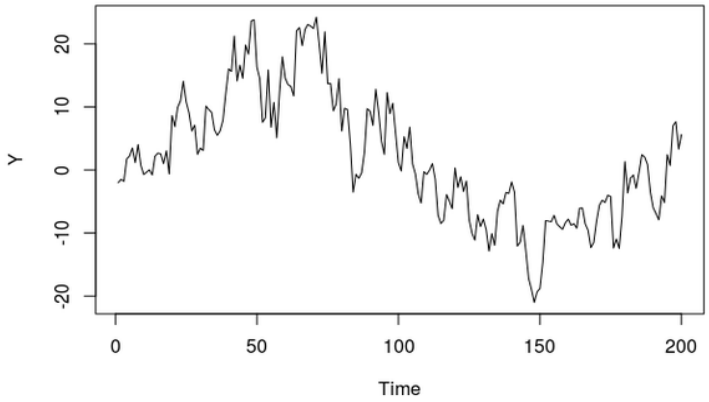

Figura 51-Esempio grafico di una serie esplosiva.

Al fine di trasformare i dati, deve essere effettuata una vera e propria integrazione, cioè si deve passare dall'osservazione di y_t all'osservazione di $\Delta y_t = y_t - y_{t-1}$.
L'integrazione della funzione porta ad una stazionarietà (Figura 52).
Il secondo modello integrato nel metodo ARIMA è il modello a media mobile che assume che il valore della risposta attuale si muova intorno ad una media costante ed è influenzato dai valori precedenti di scarto della media a cui si aggiunge uno scarto casuale.
Questo si traduce nella funzione:

$$y_t = \beta_0 + \beta_1 \cdot \varepsilon_{t-1} + \varepsilon_t$$

che rappresenta un modello MA di livello 1.

Un modello a media mobile è relativamente semplice da trattare, in quanto non è prevista la stazionarietà ma, al fine di stimare i suoi coefficienti, prevede una invertibilità. Poiché anche in questo caso gli errori sono dipendenti tra loro, si utilizzano le stime a massima verosimiglianza.
L'insieme dei modelli AR, MA e I, sopra spiegati, costituiscono il modello ARIMA.
Esso comprende dunque i tre modelli, ereditandone i presupposti, cioè la stazionarietà della serie e l'invertibilità.
È un modello relativamente semplice [14,15], ma per i suoi prerequisiti non sempre applicabile in modo ottimale.

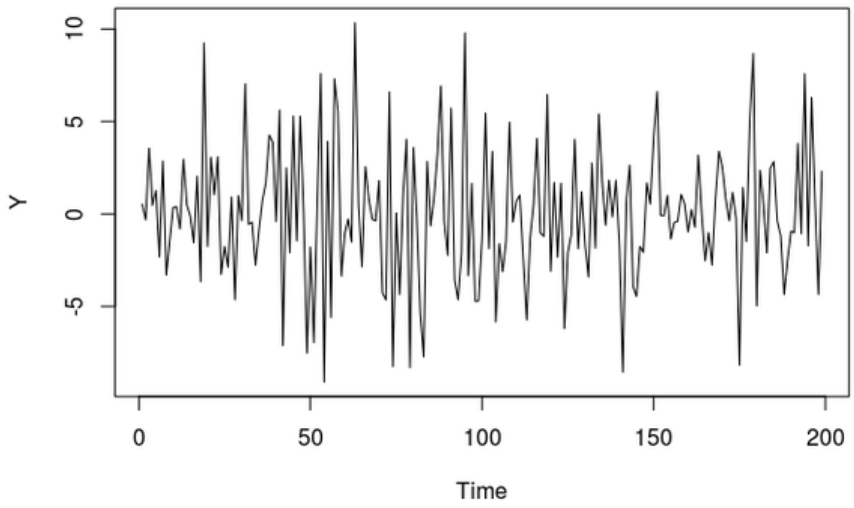

Figura 52-esempio di funzione stazionaria ottenuta tramite integrazione della funzione descritta dalla figura 51.

5.6 K-means

Il k-means [16,17]è un algoritmo di clustering anch'esso molto utilizzato, e lavora dividendo i dati in un certo numero di "clusters", cercando di minimizzare la varianza totale tra i clusters.
Per fare questo utilizza una procedura iterativa con la quale per prima cosa effettua una suddivisone in K partizioni e assegna i punti alle partizioni in modo casuale, calcolando il punto medio (centroide) di ogni cluster. Successivamente associa ogni punto al cluster il cui centroide è più vicino ad esso riconfigurando di fatto tutti cluster e ridefinendo il centroide ogni volta, calcolando la media di tutti i data points che sono assegnati al nuovo cluster.
Al termine dell'algoritmo si ha una classificazione dei punti in base ai centroidi, per cui non si ha più alcuna modifica ai clusters. Questo può voler dire che:

- La somma delle distanze è minima
- Nessun punto cambia cluster
- Si raggiunge un numero massimo di iterazioni

Il k-means è molto utile quando è possibile creare gruppi di oggetti simili e quando tali gruppi sono ben distinti e classificabili. Inoltre è ottimo in caso si conosca a priori il numero K di clusters da creare.
Di per sé è un algoritmo molto veloce, tuttavia la casualità con cui distribuisce i dati inizialmente, potrebbe portare a differenti risultati, pertanto i risultati potrebbero non essere sempre coerenti.
Dalla letteratura studiata, si evince che, soprattutto in caso di Big data, e in concomitanza con l'uso di reti neurali, questo algoritmo ha una buona capacità previsionale [18,19,20].

5.7 Suppor Vector machine

Anche le Support Vector Machine (SVM) è un algoritmo molto utilizzato nel machine learning in quanto non richiede grosse potenze di calcolo ed ha una buona accuratezza [21-22]. Esso si propone di trovare un iperpiano in uno spazio N-dimensionale, dove N è il numero delle caratteristiche legate ad una "classificazione" dei dati. Ad esempio, se abbiamo definito due classi, l'SVM assegna i nuovi dati ad una delle due classi, fungendo di fatto da classificatore lineare binario. Questo può essere esteso a più classi. Per separare le classi non esiste un solo piano, ne esistono molti ma deve essere preso quello più distante dai punti di entrambe le classi, per cui il problema si traduce nella massimizzazione dei margini al fine di ridurre la possibilità che un punto venga associato ad una classe errata.
I vettori di supporto, sono i punti che sono più vicini al piano, ed è in base alla loro posizione che viene deciso l'iperpiano migliore. L'algoritmo, dunque, in base alle classi, cerca i piani al fine di separare le classi. Se ne trova più di uno, considera quello più distante dai vettori di supporto. Se non ne trova alcuno, utilizza una mappatura non lineare per trasformare i dati in una dimensione superiore.
Questo algoritmo risulta molto utilizzato in letteratura, per lo studio di dati di consumi di edifici [23,24,25].

5.8 Reti neurali

Le reti neurali sono modelli matematici appositamente studiati per emulare il comportamento del cervello umano, cercando di avvicinarsi alle sue prestazioni cognitive.
A tal fine il cervello umano viene modellizzato come una rete, con una certa capacità di elaborazione, costituita da nodi che sono i neuroni artificiali [26-29].
L'imitazione di processi biologici, prende il nome di deep learning [30], e consiste nella creazione di più reti di neuroni, stratificate in modo da elaborare l'informazione attraverso step successivi in modo sempre più completo.
Per grandi linee una rete neurale è costituita da 3 strati,

- Strato di Input: effettua una prima elaborazione dei dati in ingresso alla rete
- Strato Hidden: effettua una elaborazione più completa secondo algoritmi definiti
- Strato di output: raccoglie i risultati dell'elaborazione e prosegue con successivi livelli. I dati in uscita da questo strato fungono da dati di input per una nuova rete.

Al fine di permettere una elaborazione dei dati, simulando il processo biologico umano, è necessario l'addestramento della rete, cioè mettere in grado la rete di capire come comportarsi a seconda dei dati che arrivano e che vengono elaborati.
Questo avviene utilizzando dei set di dati, studiati apposta per questo scopo, e che dipendono dalla specifica applicazione della rete neurale. Ad esempio, si può creare una rete neurale per il riconoscimento facciale, passando come set di dati alcune immagini di volti. La rete addestrata riuscirà così a riconoscere un soggetto a partire da una serie di immagini. Per addestrare una rete si utilizzano algoritmi di machine learning che sono suddivisi in varie categorie di apprendimento a seconda dello scopo della rete neurale. Nell'analisi predittiva le reti neurali sono molto utilizzate in quanto sono in grado di identificare ed analizzare situazioni e fenomeni anche moto complessi [31-32]. Spesso la problematica computazionale legata alle performance viene superata dalla tecnologia che offre sistemi di elaborazione più potenti ed in cambio ci offre grandi opportunità di elaborare dati ottenendo risultati molto accurati.

Entrando un po' meglio nel dettaglio, un singolo neurone possiede delle informazioni elementari che vengono memorizzate in un'unica informazione, secondo una logica di in/out (Figura 53).

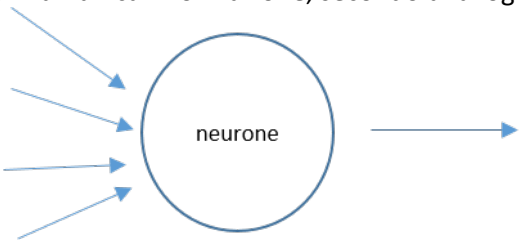

Figura 53--neurone che riceve più dati di input e restituisce un output.

Queste informazioni vengono acquisite tramite l'apprendimento, in cui coppie di dati di input e output vengono inviate alla rete come esempi. Ogni aggiornamento della coppia di dati risulterà in un aggiornamento della rete e, a mano a mano, ad un input di ingresso corrisponderà in uscita un output ottimale. Ciascun neurone acquisisce una serie di informazioni, la cui somma pesata raggiunge un livello di soglia detta "soglia di attivazione", oltre la quale il neurone si attiva e invia il suo output agli altri neuroni a lui connessi.
La funzione che descrive il processo di attivazione, prende il nome di funzione di attivazione, e può essere di varie tipologie:

- Può essere una funzione a gradino in cui gli stati sono 0 o 1, pertanto il neurone è solo attivo/non attivo (Figura 54).

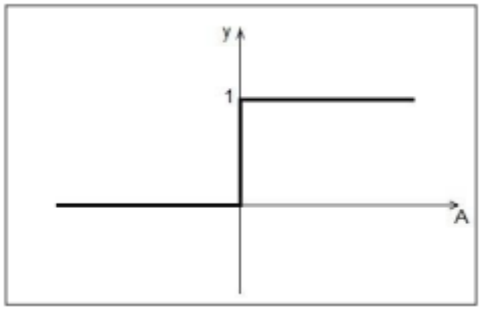

Figura 54-funzione a gradino.

- Può essere una funzione bipolare in cui gli stati sono -1 o 1, pertanto il neurone è solo attivo/non attivo (Figura 55).

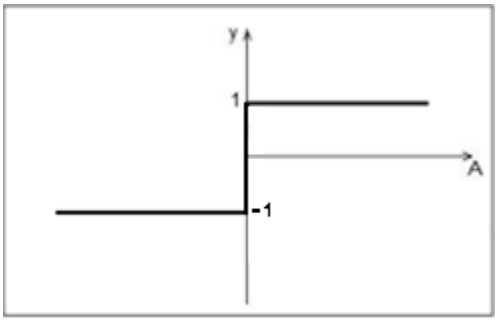

Figura 55-funzione bipolare.

- Può essere una funzione lineare, per permettere la trasmissione di segnali di varia entità (Figura 56).

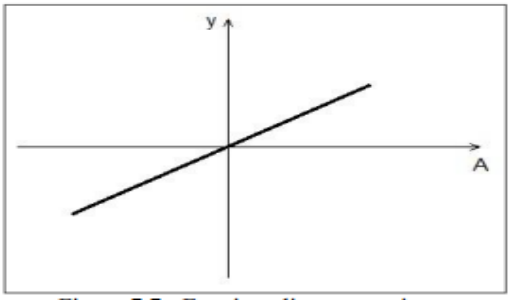

Figura 56- funzione lineare.

- Può essere una funzione sigmoidea (Figura 57).

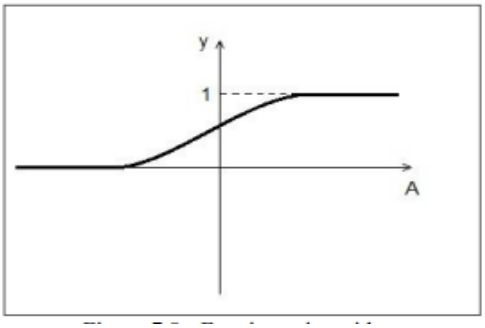

Figura 57-– funzione sigmoidea.

- Può essere una funzione tangente iperbolica (Figura 58).

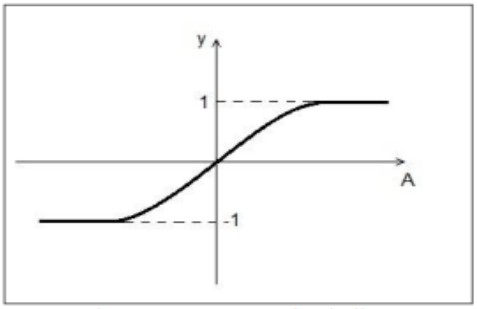

Figura 58-funzione tangente iperbolica.

È importante valutare il tipo di apprendimento per una rete neurale che può essere:

- Supervisionato: alla rete viene fornito un set di input a cui corrisponde un output. In tal modo la rete costruisce da sola la relazione tra input e output, generalizzando il processo.
- Non supervisionato: alla rete vengono forniti solo dati di input, senza alcun output, in modo che la rete possa identificare una relazione logica tra gli stessi input.

In fase di apprendimento, i pesi delle informazioni vengono definiti. A valle dell'apprendimento, c'è la fase di test in cui vengono forniti dei nuovi dati che non vanno a modificare i pesi definiti in fase di apprendimento. Questa fase è molto importante, perché permette di valutare la capacità della rete neurale di "generalizzare" il problema, rendendola particolarmente utile nello studio di fenomeni non lineari, per i quali uno studio analitico non è possibile.
In caso di apprendimento supervisionato, per la valutazione della rete si considera l'errore medio tra la risposta attesa e la risposta fornita dalla rete per ciascuna coppia input/output, valutandone l'entità in base ad un limite che si vuole raggiungere, stabilito a priori.
In caso di apprendimento non supervisionato non esiste una misura di errore, e quindi si utilizzano altri parametri o indici che quantificano la capacità di una rete di convergere verso una soluzione.
A titolo di esempio si riporta l'algoritmo di base su cui sono stati sviluppati i successivi algoritmi più moderni, cioè l'algoritmo Back-propagation.
Esso è un algoritmo di apprendimento che si basa sul confronto tra il valore in uscita dal sistema e il valore desiderato. La differenza tra i due valori, cioè l'errore, costituisce la base calcolo per i pesi sinaptici della rete che si aggiorna fino a che i valori di uscita tendono a convergere verso i valori desiderati.
Questo avviene progressivamente, attraverso una retroazione che permette di aggiornare i pesi fino ad ottenere l'obbiettivo.
In Figura 59 viene mostrato il funzionamento dell'algoritmo, in caso di rete neurale con un solo nodo N e funzione in ingresso x; ovviamente il ragionamento può essere esteso a più nodi.
y è la funzione in uscita dal nodo N, mentre y_d è la funzione desiderata. L'errore è:

$$\epsilon = y_d - y$$

e, insieme ad altri due parametri che sono il tasso d apprendimento n e il momento a, consente di ricalcolare i pesi sinaptici della rete p.

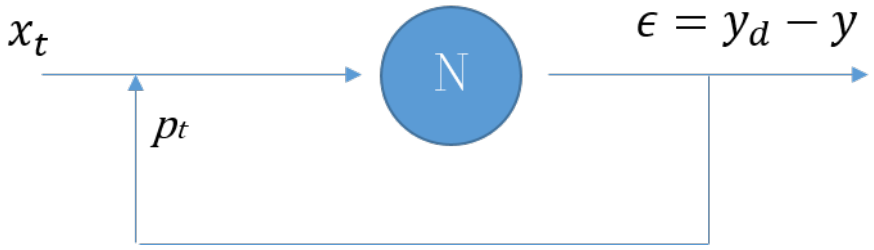

Figura 59–Schema dell'algoritmo di back propagation in caso di un solo neurone N. il valore in ingresso x , fornisce una uscita y che, sottratta all'uscita desiderata, viene utilizzata per calcolare i pesi della rete.

Il tasso di apprendimento valuta la velocità con cui la rete si addestra e varia tra un valore minimo, 0, e un valore massimo, 1. Un aumento del tasso di apprendimento rischia di aumentare l'instabilità del sistema generando oscillazioni, in quanto sono maggiori le variazioni da apporre al peso; pertanto, per ridurre questa problema si introduce il momento, sempre compreso tra 0 e 1, che aggiusta i valori dei pesi.

Nel caso di N coppie di vettori in input (x_i, y_i) con i=1,…,N si deve minimizzare la funzione di errore che considera la differenza tra il vettore di uscita y_{ij} ei il vettore y_{idj}:

$$E = \frac{1}{2}\sum_i \sum_j \left(\overline{y_{idj}} - \overline{y_{ij}}\right)^2$$

Essa dipende anche dai pesi che ad ogni retroazione vengono aggiornati fino alla convergenza, e viene minimizzata attraverso algoritmi (ad esempio discesa del gradiente), quindi viene calcolata e poi applicata la correzione da apportare ai pesi.
Un altro tipo di rete molto interessante, che a differenza della back-propagation, è auto-organizzante, cioè crea una mappa di classificazione non guidata dall'utente, è la rete di Kohonen.
In base ad un ingresso, alcuni neuroni vengono attivati; quello che presenta il vettore dei pesi più vicino all'input è il nodo vincitore e i suoi pesi vengono aggiornati per rinforzare lo stimolo, aumentando il tasso di apprendimento.
I restanti nodi adiacenti vengono invece attenuati modificando anch'essi i propri pesi.
In questo modo la mappa viene costruita considerando i neuroni vincitori per ogni set di input, ottenendo una classificazione che, ricevuto un nuovo input, associa i pattern ricorrenti in base alla mappa costruita.
Molti sono gli algoritmi di apprendimento che possono essere utilizzati, e questo rende le reti neurali molto flessibili, seppur con una relativa complessità dovuta proprio alla correttezza della fase di apprendimento.
Le reti neurali sono uno degli strumenti più utilizzati in letteratura; infatti sono strumenti flessibili e che permettono di gestire modelli più complessi [33-38].

5.9 Il DataSet

I dati attualmente in nostro possesso sono riferiti a periodi di tempo limitati, tuttavia è stato comunque possibile, in attesa di una quantità più consistente di informazioni, pensare a strumenti e metodologie di analisi che siano adeguate al nostro scenario, in particolare all'Illuminazione pubblica.
Come detto, l'illuminazione pubblica è un punto centrale di ogni realtà cittadina, in quanto è annoverata tra le strutture energivore per i suoi grandi consumi dovuti alla grande estensione degli impianti.
Sfortunatamente, il sedime storico degli impianti di illuminazione pubblica ha fatto passare questo servizio in secondo piano rispetto ad altri servizi cittadini. Infatti, da un lato gli impianti non hanno potuto avvalersi di innovazioni tecnologiche significative almeno fino al 2010, dall'altro, la loro naturale granulosità ne ha impedito un agevole censimento delle caratteristiche. Il risultato è che spesso questi impianti risultano

obsoleti ed inefficienti. D'altro canto, proprio a causa della trascuratezza da parte delle amministrazioni comunali di questi impianti, i margini di miglioramento in relazione al consumo energetico sono molto ampi anche perché corroborati dalle tecnologie oggi disponibili.
Il continuo richiamo politico alla sostenibilità ambientale e al risparmio energetico ha portato all'attuazione di importanti programmi di investimenti da parte della politica per la riqualificazione degli impianti di illuminazione pubblica e la riduzione dei consumi energetici.
Nello scenario appena delineato, il PELL IP è molto importante in quanto permette di aiutare le amministrazioni comunali nel censimento degli impianti, offrendogli anche strumenti per la riqualificazione energetica.
Una volta caricata la scheda censimento dell'impianto, in presenza di dati relativi alle grandezze energetiche provenienti dagli smart meters dei fornitori di energia elettrica, è sicuramente possibile effettuare un'analisi della domanda passata e presente di energia elettrica, e implementare algoritmi, descritti nei paragrafi precedenti, per la previsione di consumi futuri.
I dati disponibili sono relativi ai consumi energetici di impianti di illuminazione pubblica di tutta Italia; per la loro numerosità possono essere annoverati tra i cosiddetti "BigData". Infatti, se pensiamo ad esempio ad uno smart meter che fornisce misure quartorarie (un campione ogni 15 minuti), per ogni punto di consegna (Point of Delivery – POD) sono disponibili, giornalmente, 96 informazioni quartorarie sui suoi consumi. Ciò significa che per ogni POD, in un anno, sono disponibili oltre 30.000 informazioni di misura che, estese a tutti i POD formano numeri molto consistenti.
A questo si aggiunge che ciascuna delle grandezze componenti il consumo di energia (consumo di potenza attiva e reattiva, dati di corrente e di tensione, fattori di potenza, se il contatore è trifase o monofase ecc.) possono essere sfruttate per altri tipi di indagini (consumi anomali, manutenzione, ecc.). Quindi, rispetto a studi precedenti rintracciabili in letteratura, il fatto di avere a disposizione a regime una quantità di dati così vasta, ci permette di avere una granularità di informazioni notevole che rappresenta un unicum nel panorama mondiale che già da sola giustificherebbe una specifica attività di ricerca.
Questa grande quantità di informazioni disponibili, pone delle sfide tecnologiche, in quanto i dati devono essere:

- Rilevati: ciò avviene tramite gli smart meters che raccolgono i dati e li inviano verso i supporti per la memorizzazione.
- Acquisiti: il sistema di acquisizione dei dati deve essere robusto ed efficiente. Infatti è probabile che molti dati arrivino in tempi molto ravvicinati, quindi la gestione di eventuali thread deve essere flessibile e rapida onde evitare di perdere dati importanti. Questo ha un impatto da un punto di vista sistemistico (i server e i broker che si occupano di gestire i messaggi con i dati), ma anche da un punto di vista della storicizzazione (i dati devono essere salvati in modo efficiente e rapido).

  Tutto questo si traduce in sfide tecnologiche che, per nostra fortuna, sono state affrontate attraverso strumenti già predisposti allo scopo (DB noSql, Hadoop, strutture già utilizzate dai grandi colossi del web come Google o Facebook).
- Elaborati: i dati devono essere elaborati, e questo prevede un loro recupero dal database. È evidente che tale funzionalità debba essere rapida, eventualmente in tempo reale, senza problemi legati alle performances o alla carenza di memoria fisica o CPU degli elaboratori. A tale fine esistono strumenti dedicati, che coniugano efficienza ed efficacia, applicando algoritmi che permettono di utilizzare in modo appropriato le informazioni (Apache Spark, MapReduce ecc.).
- Analizzati: infine i dati devono essere analizzati e studiati attraverso algoritmi specifici che si avvalgano di pacchetti applicativi appositamente studiati (ad esempio MLib, una libreria di Apache Spark per il machine learning).

Attualmente i dati che arrivano sul sistema sono in formato JSON, e vengono inviati dai gestori verso un broker MQTT che si occupa di gestire le varie richieste.
Il JSON contiene molti dati, in cui le grandezze sono legate al tempo (al momento quartorario ma può essere orario, giornaliero, mensile ecc.).

Ciò vuol dire che, per ogni grandezza, è possibile definire una serie di valori che sono funzioni del tempo.
Alcune grandezze di interesse sono indicate di seguito:

- Il consumo di energia attiva, per monitorare eventuali inefficienze dovute, ad esempio, all'utilizzo di sorgenti luminose non adeguate, oppure per prevedere una futura richiesta di energia, anche relazionata ad altre variabili (es. dimensione dei comuni, densità di popolazione ecc.).
- Il consumo di energia reattiva, per monitorare la qualità dell'impianto e l'eventuale necessità di rifasamento.
- L'andamento della tensione per verificare eventuali cadute elevate della stessa.
- Eventuali perdite di potenza.
- Variazioni anomale della corrente per predire guasti o problemi sull'impianto.
- Analisi del fattore di potenza che possa far immaginare un rifasamento adattivo.
- Parametri legati all'affidabilità dell'impianto per una sua manutenzione programmata.

Dal momento che si deve procedere ad una analisi dei dati, e tale analisi deve considerare necessariamente delle variabili legate alla definizione di scenari, il data set costituito dai dati di consumo, deve essere integrato con i dati dell'impianto, che sono presenti nella parte relativa al PELL IP statico, cioè nella scheda censimento.
Analogamente a quanto fatto per i KPI, è necessario considerare alcuni parametri che definiscono l'impianto; ad esempio, la zona omogenea è definita come una porzione di area che presenti le stesse caratteristiche e le stesse prestazioni di illuminazione, al fine di mantenere la sicurezza di veicoli, cose e persone o per altre esigenze.
In tale contesto, per un singolo impianto che è costituito da più POD, è possibile valutare l'andamento dei consumi, ed effettuare previsioni.
Alcuni parametri relativi al comune che gestisce gli impianti possono essere presi in considerazione, come ad esempio:

- Il numero di abitanti e/o densità di popolazione.
- La superficie del comune.
- Se l'impianto è stato riqualificato già oppure no.
- Il numero dei punti luce totali e la potenza totale assorbita.

In letteratura sono descritti anche altri indicatori che possono essere utili alla predizione dei consumi/domanda di energia elettrica, e che non sono indicati nelle schede censimento come, ad esempio, la temperatura media nel comune, giornaliera, mensile o stagionale, ma anche economici e sociali [39-41]. Questi ultimi sono ovviamente più significativi in caso di ambiti residenziali, cioè abitazioni oppure edifici, mentre nel caso di illuminazione pubblica sono meno rilevanti.

5.10 Indicatori e proposta operativa

I dati provenienti dagli smart meters, sia di consumo che delle singole grandezze elettriche monitorate (corrente, tensione, fattore di potenza, ecc.), sommati ai dati anagrafici di ogni impianto, indicati nelle schede censimento e salvate sul database del PELL IP, rappresentano un database ricco di informazioni sia per granularità dell'informazione sia, a regime, per quantità.
Gli algoritmi presentati nei paragrafi precedenti, insieme ai dati ad oggi disponibili hanno avuto lo scopo di farci seguire una strada per ottenere indicatori/previsioni di consumo o efficienza energetica.
Sulla base di quanto visto abbiamo immaginato e vorremmo proporre una road map che utilizzi reti neurali che, prendendo in input alcuni parametri, permettano di valutare l'impatto degli stessi in modo completo ed efficiente.
La nostra proposta operativa può essere descritta dalla figura 60, in cui il modello presenta dei parametri di input, ciascuno corrispondente ad un neurone. Il dato in uscita dalla rete neurale corrisponde ad un solo neurone nello strato di output. All'interno del modello vengono configurati i pesi sinaptici in seguito alla fase di apprendimento della rete.

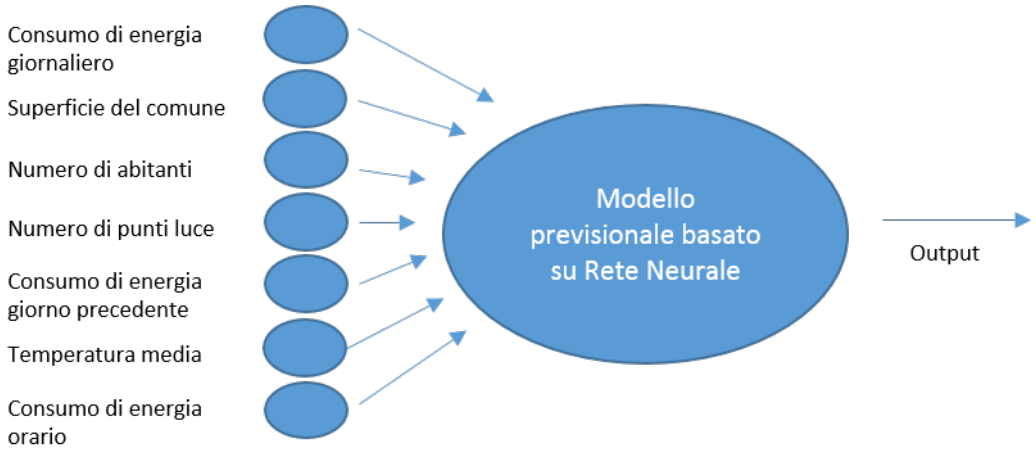

Figura 60–Schema di massima di una rete neurale artificiale.

Come input della rete neurale è sicuramente possibile considerare vari parametri che crediamo possano essere interessanti per il nostro scopo quali:

- Il consumo di energia elettrico, ricavato dai dati provenienti dagli smart meters come dato quartorario, eventualmente valutando una aggregazione per ora, per giorno,
- Dati relativi all' energia reattiva
- Dati di corrente
- Eventuali valori riferiti a periodi temporali precedenti
- Eventuali valori riferibili alla stagionalità
- Variabili meteorologiche, ad esempio la temperatura media del comune in cui è ubicato l'impianto e/o l'umidità dell'aria;
- Il numero di abitanti
- La superficie del comune ove è ubicato l'impianto
- Il numero dei punti luce

Questi tre ultimi dati sono recuperabili dalla scheda censimento attraverso opportune funzioni.
Un altro dato che si trova sulla scheda censimento, e come detto potenzialmente innovativo per l'analisi dei nostri dati, è quello relativo alla "zona omogenea", cioè un'area che necessita di uguali prestazioni illuminotecniche per quanto riguarda l'illuminazione artificiale al fine di garantire la sicurezza della circolazione veicolare o pedonale in primis o per altre esigenze.
Infatti la suddivisione dell'impianto in zone omogenee permette di avere una più ampia variabilità di casistiche e di situazioni, che possono essere ben implementate in una rete neurale.
Le reti neurali, per loro natura, non hanno un comportamento lineare e la loro topologia va verificata caso per caso affinandola sia nel numero di ingressi sia nel numero di neuroni dello strato intermedio. In via preliminare possiamo immaginare una topologia che non usi tutti gli input disponibili nel dataset. Si possono inizialmente prevedere più modelli, ciascuno con pochi input in ingresso; ad esempio, un modello con consumo di energia giornaliero, valori riferiti al giorno precedente, e dati dell'impianto. Verificatone il comportamento si potrà poi arricchire il modello prendendo in considerazione altri ingressi.
Nello strato di uscita possiamo già pensare ad un solo neurone come output, mentre, come accennato, per lo strato intermedio valuteremo in corso d'opera sia la migliore quantità di neuroni sia, eventualmente, il numero degli strati effettuando delle prove.
I dati di input, per essere utilizzabili dalle rete neurale, devono essere codificati opportunamente, ad esempio un dato orario può essere codificato attraverso un opportuno codice binario.
Per valutare le prestazioni della rete si può utilizzare il "Mean Absolute Percentage Error o MAPE" (vedi paragrafo 5.2) cioè l'errore percentuale medio tra la previsione e il dato effettivo, oppure il Mean Absolute Percentage Error o MAE cioè l'errore medio assoluto.

Già in base a questi due parametri è possibile avere una idea della bontà della previsione ed, eventualmente, affinare il modello predittivo.
Per quanto riguarda l'addestramento della rete, si può pensare ad un classico algoritmo di back propagation, in cui la rete neurale si crea una sorta di "mappatura" tra input e output, minimizzando l'errore tra l'output prodotto e l'output desiderato.
Questa minimizzazione dell'errore tra output desiderato e prodotto dalla rete, implica l'ottimizzazione dell'apprendimento, e questo processo continua fino a che la rete "converge", cioè l'errore risulta minimo.
Il processo di apprendimento della rete configura dei "pesi sinaptici", cioè i pesi associati ai collegamenti tra gli strati di neuroni.
Per implementare un modello previsionale, bisogna considerare i dati da utilizzare per l'apprendimento (training set), e quelli per testare o validare il modello stesso, oltre ai dati di input.
Per quanto riguarda i dati, l'utilizzo di una rete neurale per l'analisi non esclude un primo processamento dei dati tramite, ad esempio, metodi matematici come le regressioni, anzi una prima analisi dei dati è sempre auspicabile. Infatti essa può fornire l'idea di un andamento temporale della previsione, evidenziandone determinate caratteristiche come la periodicità o la stagionalità. Inoltre tramite una pre-analisi è possibile identificare molte situazioni che potrebbero falsare una previsione, permettendoci quindi di rimuoverle oppure, all'abbisogna, di inserirle nell'analisi. Ad esempio possono essere rilevati picchi anomali dovuti ad errori nella misurazione, oppure picchi casuali, o ancora possono esserci "buchi" nei dati dovuti a valori assenti, oppure irregolarità di cui valutarne la casualità.
A seconda del modello che si implementa, è possibile avere come output informazioni di vario tipo:

- la previsione energetica giornaliera media e massima,
- la previsione per i giorni successivi
- la previsione di eventuali guasti ad un impianto o a parte di esso.

5.11 Software per l'elaborazione dei dati

Come descritto nel paragrafo 6.9, i dati a cui abbiamo accesso sono attualmente riferiti ad un periodo di tempo limitato. Tuttavia la potenziale numerosità dei dati ci impone di valutare in modo specifico gli strumenti informatici adeguati alla situazione.
Per l'elaborazione dei big data sono disponibili molti strumenti che permettono il loro recupero e la loro trasformazione.
Esistono framework che possiedono librerie specifiche anche per il Machine Learning e il data mining (metodo di estrazione di informazioni da grandi quantità di dati) o per la gestione di modelli, e la maggior parte sono piattaforme open source disponibili gratuitamente o con licenza GPLv3 (General public license).
Da una analisi della letteratura è emerso l'uso preponderante di alcuni strumenti piuttosto che altri, ma in ogni caso la quantità di strumenti presenti per la gestione dei big data è notevole.
Infatti, nel corso degli anni c'è stata una grande evoluzione di piattaforme e, attualmente, la maggior parte dei sistemi di elaborazione dati, anche "storici", presenta evoluzioni adatte a gestirei Big Data.
Solo a titolo di esempio si può citare Matlab un pacchetto applicativo che pur risalendo agli anni '80 nella sua forma iniziale è stato costantemente aggiornato ed è molto usato nelle università e molto presente anche nella letteratura scientifica perché molto completo ed affidabile ma con il limite di essere un software a pagamento. Anche Matlab da anni è stato adeguato per gestire dei Big Data, sia attraverso librerie specifiche per il Machine Learning, sia attraverso l'interazione con database big Data come Hadoop/HDFS o database noSql, o ancora attraverso l'integrazione con tecniche di programmazione MapReduce per analizzare grandi quantità di dati anche in caso di limiti di memoria.

Apache spark (FAQ di Apache Spark), la nostra prima scelta per l'elaborazione dei dati energetici, è una libreria open source che permette di elaborare e gestire dati su larga scala in modo rapido ed efficiente.

Se da un punto di vista di memorizzazione di big data esistono piattaforme specifiche come Hadoop, da un punto di vista computazionale è necessario utilizzare strumenti che permettano di ottenere il massimo delle performance.
Questo deve necessariamente avvenire tramite l'incremento degli strumenti di calcolo, quindi computer (processori, memoria ecc.), detti nodi, che vengono aggregati in cluster, cioè in gruppi che si coordinano per l'elaborazione dei dati.
Un cluster può comprendere moltissimi nodi, anche migliaia, infatti il suo cluster più grande arriva ad utilizzare più di 8000 nodi.
Una problematica fondamentale di questi strumenti è la scalabilità, ovvero la capacità di adattamento al variare delle condizioni di elaborazione; infatti elaborare decine di GigaByte di dati non richiede sicuramente la stessa performance rispetto al calcolo di decine di TeraByte che, invece, prevede una maggior capacità computazionale.
Spark è scalabile, infatti per aumentare le performance e la capacità di calcolo basta aumentare il numero di cluster o di nodi o processori
Un'altra caratteristica che rende Spark molto flessibile, è la possibilità di usarlo su un cluster (ClusterMode) oppure standalone su un computer (ClientMode).
Esso comprende vari componenti (Figura 61) che sono:

- Il core che è il motore di base per l'elaborazione parallela e distribuita. Si occupa di integrarsi con i vari servizi di archiviazione(Hadoop), di gestire la memoria e i lavori sui clusters
- Spark SQL è il componente che permette di eseguire le query sui dati recuperati in vari formati, tra cui il JSON. Attraverso l'uso di varie funzioni (in Python, R, Scala Java), e combinando query e trasformazione di dati (RDD-resilient distributed dataset), è possibile gestire analisi molto complesse.
- MLib è la libreria di Spark che è dedicata interamente al machine learning, offrendo vari algoritmi di apprendimento automatico, valutazione dei modelli o importazione di dati
- Librerie per gestire grafici(Graphx) e flussi real time (Spark streaming)

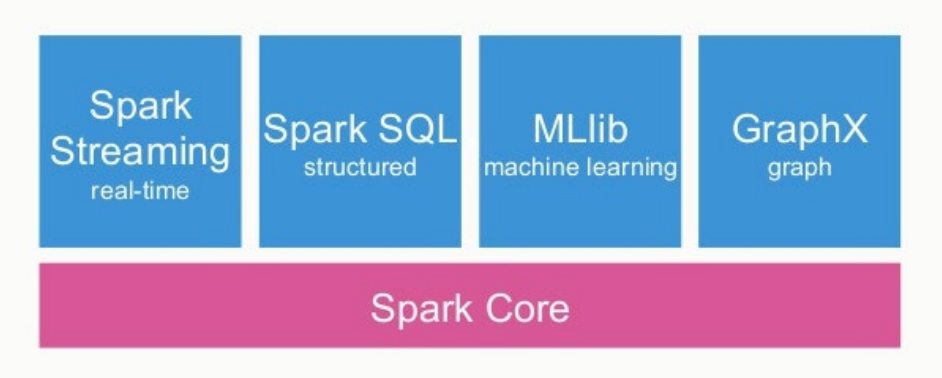

Figura 61-compoenti di Apache Spark.

In tabella 1 sono evidenziate alcune differenze tra le due applicazioni. A parte la licenza di utilizzo, entrambi gli strumenti possono essere utilizzati per effettuare analisi dati e previsioni con metodi matematici o statistici (es regressione) o librerie per il clustering (SVM, K-Means), in quanto presentano funzioni specifiche per lo scopo.
Matlab è sviluppato in C, e prevede l'utilizzo di un linguaggio proprietario per le implementazioni. A differenza di Spark, la cui libreria per il Machine learning non prevede librerie specifiche per lo sviluppo e l'addestramento di reti neurali complesse, Matlab fornisce funzioni a riga di comando ed app per creare, addestrare e simulare reti neurali anche molto complesse, nonché tools per elaborarne le performances.

SOFTWARE	Licenza	Gerstione Big Data	Librerie per metodi matematici/statistici	Librerie specifiche per Reti neurali	Librerie per clustering	Linguaggi
Apache Spark	Open source	Si	Si	No	Si	Python, R, Scala, Java
Matlab	A pagamento	Si	Si	Si	Si	Poprietario

Tabella 1 – confronto tra due software per l'elaborazione dei Big Data, Apache Spark e Matlab. Ciascuno ha delle caratteristiche che possono essere valorizzate in base all'analisi che si vuole effettuare.

6 Conclusioni

Nell'ambito del piano triennale della ricerca 2019-2021 per il sistema elettrico nazionale, per il quale l'ENEA ha predisposto il piano triennale di realizzazione, PTR 2019-2021, il Dipartimento di Scienze dell'Università degli Studi "Roma Tre" è stato interessato per una attività di ricerca dal titolo "Smart Energy in Sistemi Pubblici: Analisi di Affidabilità e Qualificazione dei Dati per Ridurre le Incertezze di Sistema".
All'interno di questo quadro generale, l'Università, durante il secondo anno (2020), è stata coinvolta nelle seguenti attività

- PELL IP,
- PELL Edifici,
- Attività informatica,
- Ricerca di algoritmi di efficienza energetica".

Per i primi tre punti, l'università ha svolto una attività di integrazione e supporto per l'analisi e la gestione di anomalie e test di nuove funzionalità relativamente all'applicazione PELL IP e per l'attività di gestione delle notifiche di funzionalità dei server attivi.
È stato inoltre fornito il supporto all'analisi e alla progettazione del nuovo portale PELL EDIFICI, anche sulla base delle problematiche riscontrate nell'attuale portale relativo al PELL IP.
Mentre le prime tre attività avevano un approccio tecnico anche se essenziale ai fini del funzionamento del sistema, la quarta ha una natura scientifica ed è stata portata avanti in modo autonomo con il fine di individuare un percorso che possa sostenere l'attività ENEA attraverso l'analisi dei dati energetici provenienti dal sistema PELL ci siamo occupati principalmente dello studio di algoritmi e tecnologie di analisi che potessero essere applicate al data set che abbiamo a disposizione.
Abbiamo quindi analizzato la letteratura individuandone tecniche e metodologie che tradizionalmente vengono adottate per l'analisi dei dati ed in particolare dei dati di natura energetica riportandone i principali e categorizzandoli in metodi matematici/statistici e metodi basati su intelligenza artificiale e machine learning, ma abbiamo poi evidenziato con un paragrafo specifico, alcuni esempi che fanno emergere le tendenze applicative più recenti o comunque che, per rigore scientifico, rappresentano riferimenti importanti da prendere in considerazione per future implementazioni.
Nell'analisi intrapresa è emerso con forza che il data set a nostra disposizione è caratterizzato da una grande quantità di informazioni, e quindi lo fa ricadere in un topic di recente affermazione che il lessico attuale inquadra con il termine di "Big Data", approccio che prevede di storicizzare i dati su framework specifici. Lo studio effettuato mostra che la struttura dei dati è nota, ed è prevalentemente costituita da serie temporali che riguardano varie grandezze elettriche, come l'energia consumata da un POD, l'intensità di corrente, la potenza attiva o reattiva ecc.
Questo pone varie problematiche:

- La scelta di uno strumento adeguato di elaborazione dei dati
- La valutazione attenta degli algoritmi da utilizzare per l'analisi dei dati.

In merito al primo punto, l'utilizzo di Hadoop per la memorizzazione dei dati di consumo fornisce un primo punto di partenza per la scelta di software da utilizzare. Oggi molte applicazioni sono interfacciabili con Hadoop, e se ci orientiamo verso una scelta di tipo open source, con disponibilità di una buona documentazione e disponibilità di librerie/pacchetti per il machine learning, una soluzione buona è l'utilizzo di Apache Spark.
In merito al secondo punto, sulla base dei dati che sono attualmente disponibili (in tipologia e numerosità), e sullo studio della letteratura in merito, sicuramente è possibile cominciare a lavorare sui metodi che gestiscano delle serie temporali, quindi modelli di regressione e modelli ARIMA.
Questo ci dà l'opportunità di approfondire il significato dei dati in nostro possesso, valutando anche la gestione di eventuali dati mancanti o errati, e di comprendere meglio le componenti effettivamente casuali e no del data set.
Anche le reti neurali presentano interessanti spunti di utilizzo, in particolare pensando a previsioni di consumo oppure alla domanda di energia, soprattutto quando la numerosità dei dati permetterà di

predisporre algoritmi di apprendimento adeguati alle esigenze di una rete neurale. Più informazioni sono disponibili, migliore è l'apprendimento da parte della rete, e migliore è la risposta ottenuta.
Il nostro Dataset si presta bene alla definizione dei consumi energetici mensili o annuali di illuminazione pubblica per comune, in base alla popolazione e alle temperature medie (mensili o annuali). Può anche, successivamente, essere esteso alle regioni oppure alla suddivisione in zona nord, centro e sud.
L'impostazione data, ci consente di farci immaginare che, con i dati attuali in nostro possesso, si possa realizzare l'analisi predittiva dei consumi di energia per gli impianti di illuminazione pubblica con modello su serie storica e poi su rete neurale. Per quest'ultima situazione si possono considerare alcuni indici in input, come ad esempio il numero dei punti luce e la potenza totale dell'impianto, dati reperibili dalla scheda censimento.

7 Riferimenti bibliografici

I riferimenti bibliografici devono essere richiamati nel testo con numeri progressivi tra parentesi quadre e riportati a fine testo con il seguente formato:

1. ENEA "*Progetto Lumiere".* Disponibile on line al sito: Progetto lumiere
2. ENEA "*PELL - Lumière & Public Energy Living Lab (PELL) per una gestione efficiente della Pubblica Illuminazione*" Disponibile al sito: PELL - Lumière & Public Energy Living Lab (PELL) per una gestione efficiente della Pubblica Illuminazione.
3. Specifiche di contenuto di riferimento PELL - illuminazione pubblica Specifica PELL IP, 23-07-2018
4. Tesi di dottorato Edoardo Moreci, https://iris.unipa.it/retrieve/handle/10447/162636/261304/Tesi%20di%20dottorato%20Ing.%20Edoardo%20Moreci%20-%20Dottorato%20in%20Energia%20%E2%80%93%20Ciclo%20XXVI.pdf
5. Box, George E. P., Gwilym M. Jenkins, and Gregory C. Reinsel. Time Series Analysis : Forecasting and Control. 4th, Wiley, 2008
6. Hamilton, J., Time Series Analysis, Princeton University Press,1994, ISBN 978-0-691-04289-3
7. W. A. Woodward, H. L. Gray, and A. C. Elliot, Applied Time Series Analysis, CRC Press, 2012.
8. Estela Bee Dagum, Analisi delle serie storiche - modellistica, previsione e scomposizione, Milano, Springer Verlag, 2002. ISBN 88-470-0146-3
9. Azadeh, Ali & Tarverdian, S.. (2007). Integration of genetic algorithm, computer simulation and design of experiments for forecasting electrical energy consumption. Energy Policy. 35. 5229-5241. 10.1016/j.enpol.2007.04.020.
10. Shumway, Robert H., Stoffer, David S. Time Series Analysis and Its Applications-With R Examples. Springer International Publishing, 2017, ISBN 978-3-319-52452-8
11. Shumway, R. H. , Applied statistical time series analysis. Prentice Hall. 1988 ISBN 978-0130415004.
12. Percival, Donald B.; Walden, Andrew T. (1993). Spectral Analysis for Physical Applications. Cambridge University Press.
13. Pandit, Sudhakar M.; Wu, Shien-Ming (1983). Time Series and System Analysis with Applications. John Wiley & Sons.
14. Rosa, José & Soares, Gustavo & Machado-Coelho, Thiago & Mendes, Marcus & Libório, Matheus & Machado, Alexei & Ekel, Petr. (2020). Previsão de Demanda de Energia Elétrica Utilizando Modelos Lineares. 10.48011/asba.v2i1.1695.
15. Tso, Geoffrey & Yau, Kelvin. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. Energy. 32. 1761-1768. 10.1016/j.energy.2006.11.010.
16. MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. 1. University of California Press. pp. 281–297. MR 0214227. Zbl 0214.46201. Retrieved 2009-04-07.
17. Hamerly, Greg; Elkan, Charles (2004). "Learning the k in k-means" (PDF). Advances in Neural Information Processing Systems. 16: 281.
18. Pérez-Chacón, Rubén & Cortés, Gualberto & Martínez-Álvarez, Francisco & Troncoso, Alicia. (2020). Big data time series forecasting based on pattern sequence similarity and its application to the electricity demand. Information Sciences. 540. 10.1016/j.ins.2020.06.014.

19. Pérez-Chacón, R.; Luna-Romera, J.M.; Troncoso, A.; Martínez-Álvarez, F.; Riquelme, J.C. Big Data Analytics for Discovering Electricity Consumption Patterns in Smart Cities. Energies 2018, 11, 683. https://doi.org/10.3390/en11030683

20. Martínez-Álvarez F., Troncoso A., Riquelme J.C., Riquelme J.M. (2007) Partitioning-Clustering Techniques Applied to the Electricity Price Time Series. In: Yin H., Tino P., Corchado E., Byrne W., Yao X. (eds) Intelligent Data Engineering and Automated Learning - IDEAL 2007. IDEAL 2007. Lecture Notes in Computer Science, vol 4881. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-77226-2_9

21. Stuart Russell e Peter Norvig, Intelligenza artificiale: un approccio moderno, Prentice Hall, 2003, ISBN 88-7192-229-8

22. James, Gareth; Witten, Daniela; Hastie, Trevor; Tibshirani, Robert (2013). "Support Vector Machines" (PDF). An Introduction to Statistical Learning : with Applications in R. New York: Springer. pp. 337–372. ISBN 978-1-4614-7137-0.

23. Grolinger, M. A. M. Capretz and L. Seewald, "Energy Consumption Prediction with Big Data: Balancing Prediction Accuracy and Computational Resources," 2016 IEEE International Congress on Big Data (BigData Congress), San Francisco, CA, USA, 2016, pp. 157-164, doi: 10.1109/BigDataCongress.2016.27.

24. Mel Keytingan M. Shapi, Nor Azuana Ramli, Lilik J. Awalin,Energy consumption prediction by using machine learning for smart building: Case study in Malaysia,Developments in the Built Environment,Volume 5,2021,100037,ISSN 2666-1659,https://doi.org/10.1016/j.dibe.2020.100037.

25. Puspita, Verlly & Ermatita, Ermatita. (2019). Time Series Forecasting for Electricity Consumption using Kernel Principal Component Analysis (kPCA) and Support Vector Machine (SVM). Journal of Physics: Conference Series. 1196. 012073. 10.1088/1742-6596/1196/1/012073.

26. Kantz, Holger; Thomas, Schreiber (2004). Nonlinear Time Series Analysis. London: Cambridge University Press. ISBN 978-0521529020.

27. Floreano D., Mattiussi C., Manuale sulle reti neurali, Bologna, Il Mulino, 2002. ISBN 978-88-15-08504-7

28. Battaglia F. Metodi di previsione statistica Springer Verlag 2007 ISBN 8847006023

29. Bishop, C. M. (1995). Neural networks for pattern recognition. Oxford: Clarendon Press.

30. Schmidhuber, J. (2015). "Deep Learning in Neural Networks: An Overview". Neural Networks.

31. Hassoun, Mohamad H. (1995). Fundamentals of Artificial Neural Networks. MIT Press. p. 48. ISBN 978-0-262-08239-6.

32. Bishop, Christopher M. (2006). Pattern Recognition and Machine Learning (PDF). Springer. ISBN 978-0-387-31073-2.

33. M. Beccali, M. Cellura, V. Lo Brano, A. Marvuglia, Short-term prediction of household electricity consumption: Assessing weather sensitivity in a Mediterranean area, Renewable and Sustainable Energy Reviews, Volume 12, Issue 8, 2008, Pages 2040-2065, ISSN 1364-0321, https://doi.org/10.1016/j.rser.2007.04.010.

34. H. M. Safhi, B. Frikh and B. Ouhbi, Energy load forecasting in big data context, 2020 5th International Conference on Renewable Energies for Developing Countries (REDEC), Marrakech, Morocco, 2020, pp. 1-6, doi: 10.1109/REDEC49234.2020.9163901.

35. A. Azadeh, S. F. Ghaderi, S. Tarverdian and M. Saberi, "Integration of Artificial Neural Networks and Genetic Algorithm to Predict Electrical Energy consumption," IECON 2006 - 32nd Annual Conference on IEEE Industrial Electronics, Paris, France, 2006, pp. 2552-2557, doi: 10.1109/IECON.2006.348098.

36. Lee, Seunghui & Jung, Sungwon & Lee, Jaewook. (2019). Prediction Model Based on an Artificial Neural Network for User-Based Building Energy Consumption in South Korea. Energies. 12. 608. 10.3390/en12040608.

37. Touzani, Samir & Ravache, Baptiste & Crowe, Eliot & Granderson, Jessica. (2019). Statistical Change Detection of Building Energy Consumption: Applications to Savings Estimation. Energy and Buildings. 185. 10.1016/j.enbuild.2018.12.020.

38. Cheng, Yao & Xu, Chang & Mashima, Daisuke & Biswas, Partha & Chipurupalli, Geetanjali & Zhou, Bin & Wu, Yongdong. (2020). PowerNet: A Smart Energy Forecasting Architecture based on Neural Networks. IET Smart Cities. 2. 10.1049/iet-smc.2020.0003.

39. Mayer, Audrey. (2008). Strengths and Weaknesses of Common Sustainability Indices for Multidimensional Systems. Environment international. 34. 277-91. 10.1016/j.envint.2007.09.004.

40. Kadoshin, Shiro & Nishiyama, Takashi & Ito, Toshihide. (2000). The trend in current and near future energy consumption from a statistical perspective. Applied Energy. 67. 407-417. 10.1016/S0306-2619(00)00033-7

41. Mardani, Abbas & Streimikiene, Dalia & Nilashi, Mehrbakhsh & Aranda, Daniel & Loganathan, Nanthakumar & Jusoh, Ahmad. (2018). Energy Consumption, Economic Growth, and CO2 Emissions in G20 Countries: Application of Adaptive Neuro-Fuzzy Inference System. Energies. 11. 2771. 10.3390/en11102771.

8 Abbreviazioni ed acronimi

ENEA: Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile
ENEL: Ente nazionale per l'energia elettrica
KPI: Key Performace indicator
kW: kilowatt
LED: Light Emitting Diode
PELL: Progetto Public Energy Living Lab
IP: illuminazione pubblica
MVC: Model View Controller
POD: Point Of Delivery
WEB: web browser
API: Application Programming interface
PHP: Hypertext preprocessor
Ajax: Asynchronous javascript and XML
JSON: Javascript object notation
PUN: prezzo unico dell'energia elettrica
HDFS: Hadoop distributed file system
REST: representational state transfer
XSD: XML schema definition
XML: extensible markup language
HTTP: hypertext transfer protocol
RDBMS: relational database management system
ER DIAGRAM: entity relation diagram
MQTT: Message Queuing Telemetry Transport

Appendice: Laboratorio di Misure Elettriche ed Elettroniche dell'Università degli Studi "Roma Tre": Curriculum Scientifico

Responsabile: Dott. Ing. Ph.D. RTI Fabio Leccese
Collaboratori: Dott. Enrico Petritoli (Assegnista di Ricerca), Dott.sa Mariagrazia Leccisi (Borsista)

Il laboratorio fa parte del Gruppo Nazionale delle Misure Elettriche ed Elettroniche (GMEE) i cui scopi principali sono lo studio delle misure o "metrologia", l'analisi di qualità fisiche e la realizzazione di campioni di misura con particolare attenzione allo studio dell'incertezza di misura.

In questo quadro generale, il nostro laboratorio segue da anni diverse linee di ricerca tra le quali la qualità dell'energia (power quality – dal 2004), l'analisi informativa dei segnali (dal 2002), i controlli di apparati locali e remoti ed in particolare di sistemi di risparmio energetico applicati ad illuminazione e riscaldamento (dal 2008), la sensoristica distribuita incluse le Wireless Sensor Network (dal 2008) e le analisi affidabilistiche di sistemi complessi (dal 2013) trovano ampia utilità e complementarità con le attività svolte in ENEA dal gruppo del Laboratorio Smart Cities and Communities

Ciascuna linea presenta peculiarità proprie che coinvolgono non solo il campo specifico delle misure, ma anche settori ad esso correlati quali l'elettronica, l'elettrotecnica, le telecomunicazioni, l'informatica e l'automazione. Il Laboratorio progetta e sviluppa sistemi di misura avvalendosi dei software più moderni come Orcad o Protel e programmando microcontrollori di varie famiglie come Microchip o Siemens, processori ARM, avendo confidenza anche con la progettazione di FPGA. I linguaggi di programmazione più usati sono il C, la piattaforma .NET e vari linguaggi "WEB oriented".

Nel campo della Didattica abbiamo nel tempo sviluppato dei percorsi all'interno dei nostri Dipartimenti rivolti al mondo dell'ambiente e dell'energia con materie come Qualità Ambientale, Qualità dell'Energia, Elementi di Misure per l'Analisi Ambientale, Alimentazione da Fonti Rinnovabili e Strumentazione Avanzata di Misura che, nel tempo, sono state apprezzate da un numero crescente di studenti.

A testimonianza dell'esperienza maturata, il nostro lavoro ha portato ad oltre 200 pubblicazioni scientifiche per la maggior parte presentate su Riviste Scientifiche Internazionali o su Congressi Scientifici Internazionali, come riscontrabile dalle 143 riportate sul Database Scopus. Di queste, 7 sono state scritte insieme ad ENEA e sempre con ENEA sono stati affrontati insieme 7 progetti di ricerca.

Si segnalano i premi nazionali/internazionali vinti dal Laboratorio

1) 2016 – PUBLONS - The sentinels of Science Awards 2016 The top 10 percent of reviewers- Certified Sentinel of Science award recipient: As one of the top 10 per cent of researchers contributing to the peer review of the field of Chemistry
2) 2018 – II Forum Nazionale Delle Misure – Sezione GMEE – Padova, 17-19 Settembre 2018: Miglior Poster per l'articolo: "Measurements of Q factor in microwave resonators: relevance of the calibration" a cura di K. Torokhtii, A. Alimenti, N. Pompeo, F. Leccese, F. Orsini, A. Scorza, S.A. Sciuto, E. Silva.
3) 2018 – IEEE International Workshop on Metrology for the Sea, October 08-10, Bari, Italy: Miglior Demo per il drone di nuova concezione con movimentazione a pendolo vincolato a cura di Eduardo De Francesco e Fabio Leccese.
4) 2019 - WEB OF SCIENCE – PUBLONS -TOP PEER REVIEWER 2019 -For placing in the top 1% of reviewers in Cross-Field on Publons global reviewer database.

Elenco di partecipazioni a progetti scientifici

Progetti Internazionali:

1) "PROGETTO DI GRANDE RILEVANZA ITALIA - SERBIA 2016-2018 sul tema di Agriculture and Food Technologies dal titolo **SMART MONITORING OF PESTICIDES IN FARMING AREAS**" Finanziato dal

Ministero degli Affari Esteri e della Cooperazione Internazionale. **Ruolo: Responsabile Scientifico**. Durata 3 anni.

Progetti Nazionali:

2) Bando PROGRAMMI DI RICERCA SCIENTIFICA DI RILEVANTE INTERESSE NAZIONALE RICHIESTA DI COFINANZIAMENTO **PRIN 2010-2011** dal titolo: "**Interazione fra minerali e biosfera: conseguenze per l'ambiente e la salute umana**"- sottosezione "**Emissioni antropogeniche di CO2: immobilizzazione per carbonatazione e discriminazione isotopica della componente fossile e non fossile**". PRIN 2010-2011, Area 04, Durata 36 mesi, Protocollo 2010 MKHT9B_007
3) Progetto **Co-Research POR FESR LAZIO 2007-2013 – Titolo SIMPLFEX** Progetti di R&S in collaborazione presentati dalle PMI del Lazio con Numero di protocollo assegnato: FILAS-CR-2011-1076 dal 09/01/2012 al 08/01/2014. **Ruolo: Responsabile Scientifico di Sede.** Durata 2 anni.
4) Progetto di ricerca: **"Sviluppo e implementazione di algoritmi per applicazioni di Smart Lighting"** per conto di ENEA – Roma, 2014. **Ruolo: Responsabile Scientifico**. Durata 1 anno.
5) Progetto di ricerca: **"Sviluppo e implementazione di indicatori di prestazione e diagnostica energetica per impianti di illuminazione pubblica"** per conto di ENEA – Roma, 2014. **Ruolo: Responsabile Scientifico**. Durata 5 mesi.

I seguenti progetti sono stati sviluppati all'interno del piano Piano Triennale della Ricerca nell'ambito del Sistema Elettrico Nazionale 2015-2017 finanziato dal Ministero dello Sviluppo Economico (MiSE) e gestito da ENEA all'interno dell'Accordo di Programma MiSE-ENEA 2015-2017.

6) **Progettazione e sviluppo prototipale di strumenti per la gestione del PELL**, per conto di ENEA – Roma, 2015. **Ruolo: Responsabile Scientifico di Sede**. Durata 5 mesi.
7) **Analisi di affidabilità e analisi dei guasti e delle criticità (FMECA) del sistema smart street**, per conto di ENEA – Roma, 2016. **Ruolo: Responsabile Scientifico di Sede**. Durata 5 mesi.
8) **Studio affidabilistico dei componenti di una linea di illuminazione "smart" stradale pubblica operativa in contesto urbano: vantaggi e criticità**, per conto di ENEA – Roma, 2017. **Ruolo: Responsabile Scientifico di Sede**. Durata 5 mesi.
9) **Studio affidabilistico preliminare dei componenti fondamentali del sistema di termoregolazione dell'edificio F-40 ENEA (Casaccia): vantaggi e criticità**, per conto di ENEA – Roma, 2018. **Ruolo di Responsabile Scientifico di Sede**. Durata 3 mesi.
10) **Smart Energy in Sistemi Pubblici: analisi di affidabilità e qualificazione dei dati per ridurre le incertezze di sistema,** per conto di ENEA – Roma, 2019-2021. **Ruolo di Responsabile Scientifico di Sede**. Durata 3 anni.

Progetti Conto Terzi:

11) Progetto di ricerca: **"Studio delle criticità delle PowerLine su Navi da guerra"** per conto della Se.Te.L. group di Roma, 2012. **Ruolo: Responsabile Scientifico**. Durata 1 mese.
12) Progetto di ricerca: **"Valutazione del Processo di Rivitalizzazione degli Accumulatori al Piombo-Acido e del Relativo Liquido Additivo"** per conto della Battery Equalizer Italia s.r.l. di Fiumicino, 2012, **Ruolo: Responsabile Scientifico.** Durata 3 mesi.
13) Progetto di ricerca: **"Evoluzioni del supporto logistico delle power line di unità navali"** per conto della Se.Te.L. group di Roma, 2013. **Ruolo: Responsabile Scientifico**. Durata 1 mese.
14) Progetto di ricerca: **"Sistema di gestione delle telecamere di guida a bordo del Rover SETEL"** per conto della Se.Te.L. group di Roma, 2020-2021. **Ruolo: Responsabile Scientifico**. Durata 7 mesi.