

# Decentralized Smart Grid Control

Christian Arlenghi 843370, Daniele Gamberoni 837236,  
Giacomo Morzenti 837395, Luca Paoletti 837161

Sarebbe un peccato arrivare ad un futuro non molto lontano in cui non potremo utilizzare energia rinnovabile a causa di reti elettriche ancora obsolete. A questo proposito abbiamo deciso di analizzare una delle possibili soluzioni a questo problema e come vedremo il metodo utilizzato richiede un ampio utilizzo del data mining e machine learning. Scopriremo che la soluzione non sarà quella di rinnovare il sistema di reti elettriche ma di utilizzarlo in maniera più smart.

## INTRODUZIONE

La modalità di produzione dell'energia elettrica sino ad un decennio fa è avvenuta in maniera totalmente centralizzata: un numero limitato di grandi centrali elettriche, prevalentemente alimentate a fonti fossili, producevano grandi quantità di energia elettrica che poi venivano trasmesse, attraverso la rete elettrica, a milioni di utenti e consumatori. Oggi governi e istituzioni stanno cercando di favorire la produzione di elettricità a partire da fonti rinnovabili non inquinanti come eolico e solare, il cui modello è, però, estremamente **decentralizzato**: oggi in Italia sono presenti circa un milione di impianti da fonti rinnovabili, installati praticamente in ognuno dei migliaia di comuni italiani. È evidente che questa **diffusione estremamente parcellizzata** delle fonti pulite può mettere a dura prova le reti elettriche, che devono essere in grado di accogliere l'elettricità prodotta da una molteplicità di punti. A complicare il quadro c'è poi l'**intermittenza della generazione rinnovabile**: eolico e fotovoltaico producono elettricità soltanto in presenza della disponibilità della fonte.

Quindi le reti elettriche devono oggi essere abbastanza **flessibili** per riuscire ad accogliere eventuali picchi di produzione rinnovabile (ad esempio nelle ore centrali dei mesi estivi) e, al contempo, assicurare che l'intero sistema elettrico non collassi nei momenti di minore generazione. Esistono soltanto due modalità per gestire la situazione: la prima è quella di aumentare la capacità della rete esistente (principalmente la rete di distribuzione, quindi dotandosi di un numero superiore di cavi e trasformatori in media e bassa tensione). L'alternativa è quella di utilizzare la capacità della rete esistente in modo più efficiente, più intelligente, più "Smart", grazie allo sviluppo di una **Smart Grid**.

Un esempio di modello di smart grid è il D.S.G.C. (Decentralized Smart Grid Control), il quale pone soluzione a due problemi principali: la stabilità delle reti elettriche e la determinazione del prezzo dell'energia consumata.

Il modello si divide in due parti: la prima, descrive fisicamente la dinamica (basata sulle equazioni di moto) dei generatori e dei carichi di energia. Un sistema elettrico è in equilibrio se vale la seguente equazione:

$$P_{source} = P_{accumulated} + P_{dissipated} + P_{transmitted}$$

La seconda parte è una sovrastruttura economica che lega il prezzo dell'elettricità alla frequenza della rete elettrica. Un'assunzione del modello è che il consumo e la produzione di energia si adatta ai cambiamenti di prezzo.

In questa relazione studieremo la stabilità di una rete elettrica, composta da un produttore e tre partecipanti (un esempio di partecipante possono essere i consumatori di energia privati), analizzando tre variabili input:

-Tau ( $t$ ): tempo di reazione dei partecipanti, delay tra cambiamento del prezzo e adattamento del consumo

-p ( $p$ ): potenza nominale consumata (negativa) o prodotta (positivo)

-Gamma ( $g$ ): elasticità del consumo al prezzo

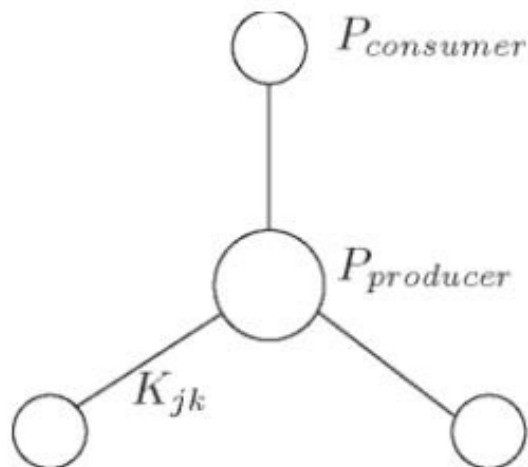
Ipotizziamo costanti tre variabili:

-tempo medio per misurare la segnalazione del prezzo (2 sec)

-forza di accoppiamento, proporzionale alla capacità della linea ( $8 \text{ sec}^{-2}$ )

-costante di dissipazione ( $0.1 \text{ sec}^{-1}$ )

A seconda dei valori che assumeranno le variabili nel produttore e nei partecipanti si stabilirà se la rete elettrica sia stabile o meno (stabf).



## MATERIALI

### COMPOSIZIONE DEL DATASET

Il dataset da noi analizzato è composto da 10000 osservazioni e 14 variabili.

13 numeric: tau1, tau2, tau3, tau4, p1, p2, p3, p4, g1, g2, g3, g4 e stab

1 factor: stabf con due fattori, stable (quando stab è negativo) e unstable (stab positivo)

### SPLITTING DEI DATI

Il dataset è stato suddiviso in training, validation e test set. L'80% dei dati è stato dedicato al training set e il rimanente 20% al test set. Il training set è stato suddiviso a sua volta in: subtraining (65%) e validation (35%).

Verifichiamo che le proporzioni della variabile risposta siano le stesse nei diversi set e coincidano con quelle del dataset completo.

PROPORZIONE STAB	Stabf=0	Stabf=1
Complete Dataset	0.36	0.64
Sub-Training	0.35	0.65
Validation	0.38	0.62
Test	0.36	0.64

### DEFINIZIONE DELLE VARIABILI

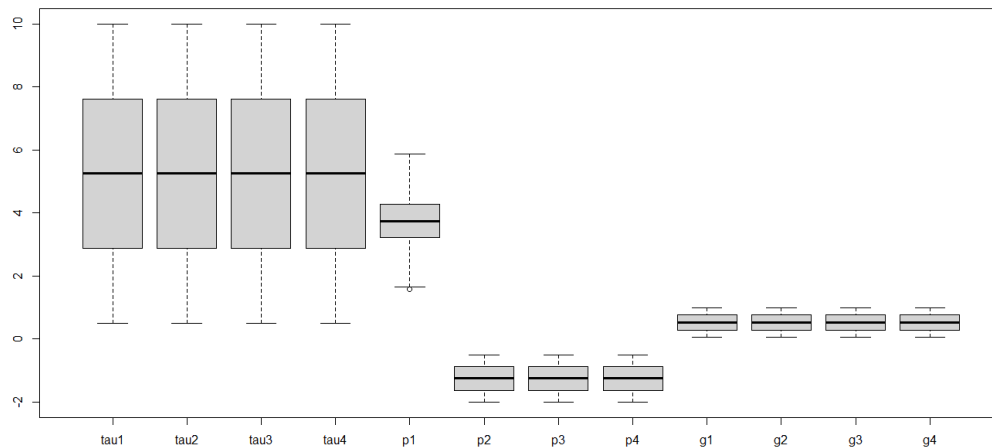
Abbiamo deciso di rimuovere la variabile "stab" dal dataset in quanto ha una perfetta correlazione con la variabile risposta, quindi toglie la significatività alle rimanenti variabili.

La variabile "stabf" è una factor a due fattori: stable quando "stab" è negativo e unstable quando "stab" è positivo, dunque c'è perfetta correlazione tra le due variabili. Abbiamo così deciso di rimuoverla per poter studiare l'effetto che le altre covariate hanno sulla risposta senza che "stab" influisca sulla loro significatività.

Inoltre impostiamo stabf=1 quando il sistema è stable e stabf=0 quando è unstable.

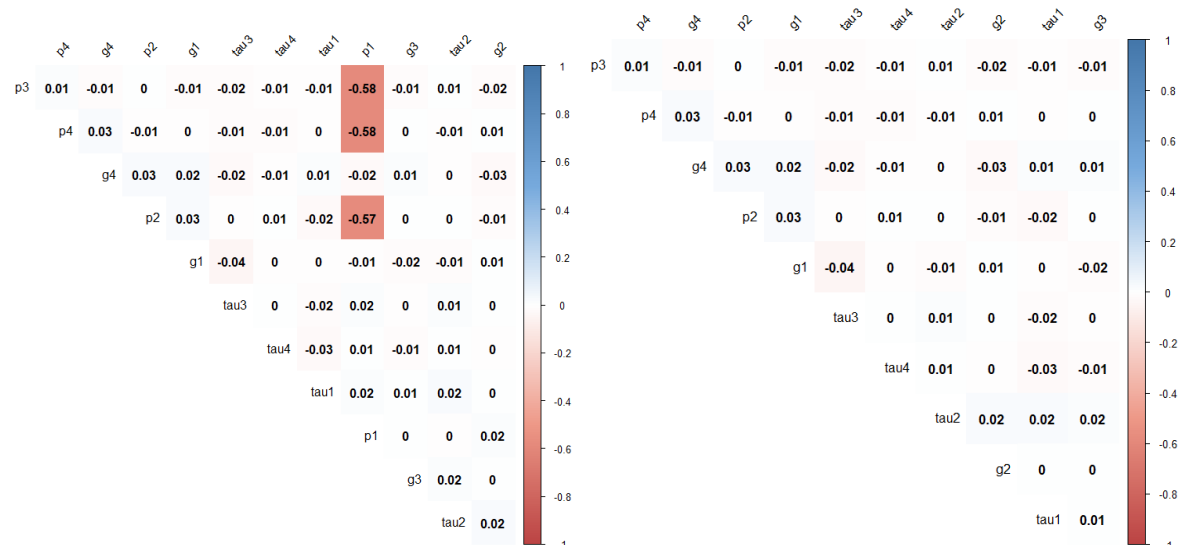
Nel dataset non sono presenti missing values, dunque non occorre eliminare nessuna osservazione né variabile.

## BOX-PLOT DELLE VARIABILI ESPLICATIVE



Da come si denota dal grafico qui sopra, le variabili esplicative non presentano valori anomali o outliers, tuttavia optiamo per una standardizzazione dei dati affinché variabili con unità di misura differenti possano essere più facilmente confrontabili.

## CONFRONTO CORRELAZIONE TRA ESPLICATIVE

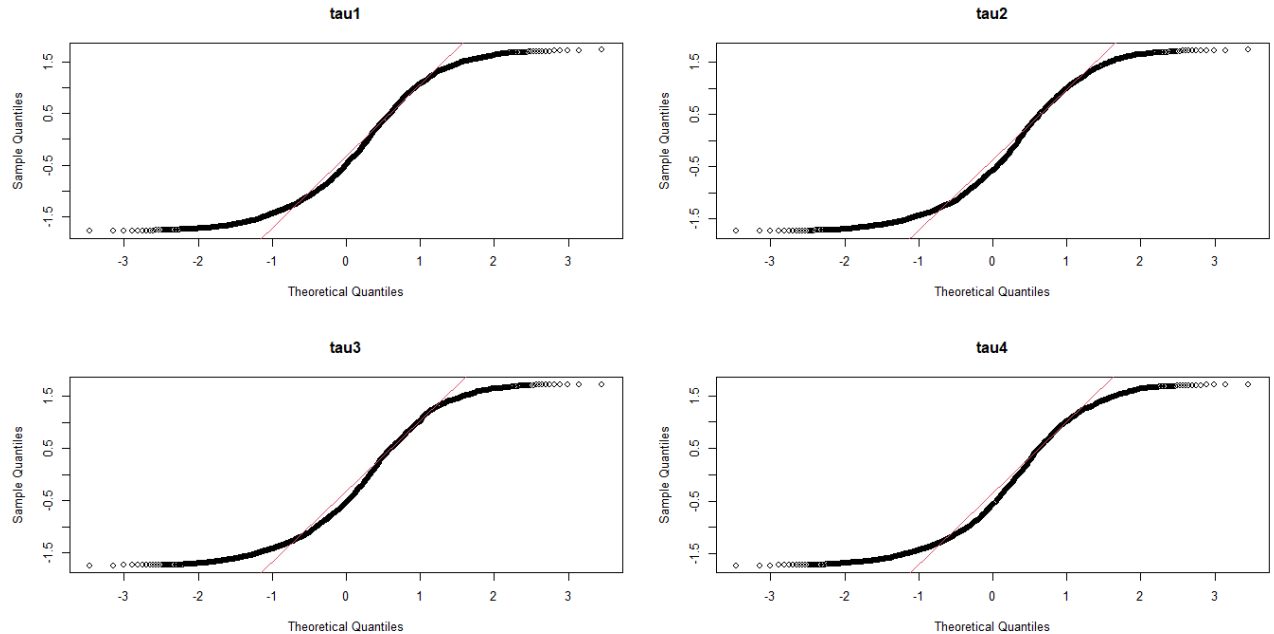


Abbiamo analizzato la correlazione tra le variabili esplicative e, come ci aspettavamo, abbiamo ottenuto un'alta correlazione tra la variabile p1 e le variabili p2, p3 e p4, in quanto p1 è combinazione delle altre variabili ( $p1 = |p2 + p3 + p4|$ ). Abbiamo quindi deciso di escludere la variabile p1 per evitare di avere correlazione tra le variabili e rischiare di avere multicollinearità.

## DISTRIBUZIONE EMPIRICA DELLE COVARIATE

Svolgiamo una prima verifica qualitativa della normalità delle covariate condizionate alla classe di appartenenza (stable, unstable) attraverso il metodo grafico che ci permette di confrontare quantile teorico della normale standard (avendo in precedenza standardizzato le variabili) e quantile empirico (Q-Q plot)

E' riportato di seguito il grafico delle variabili tau condizionate alla classe stable:

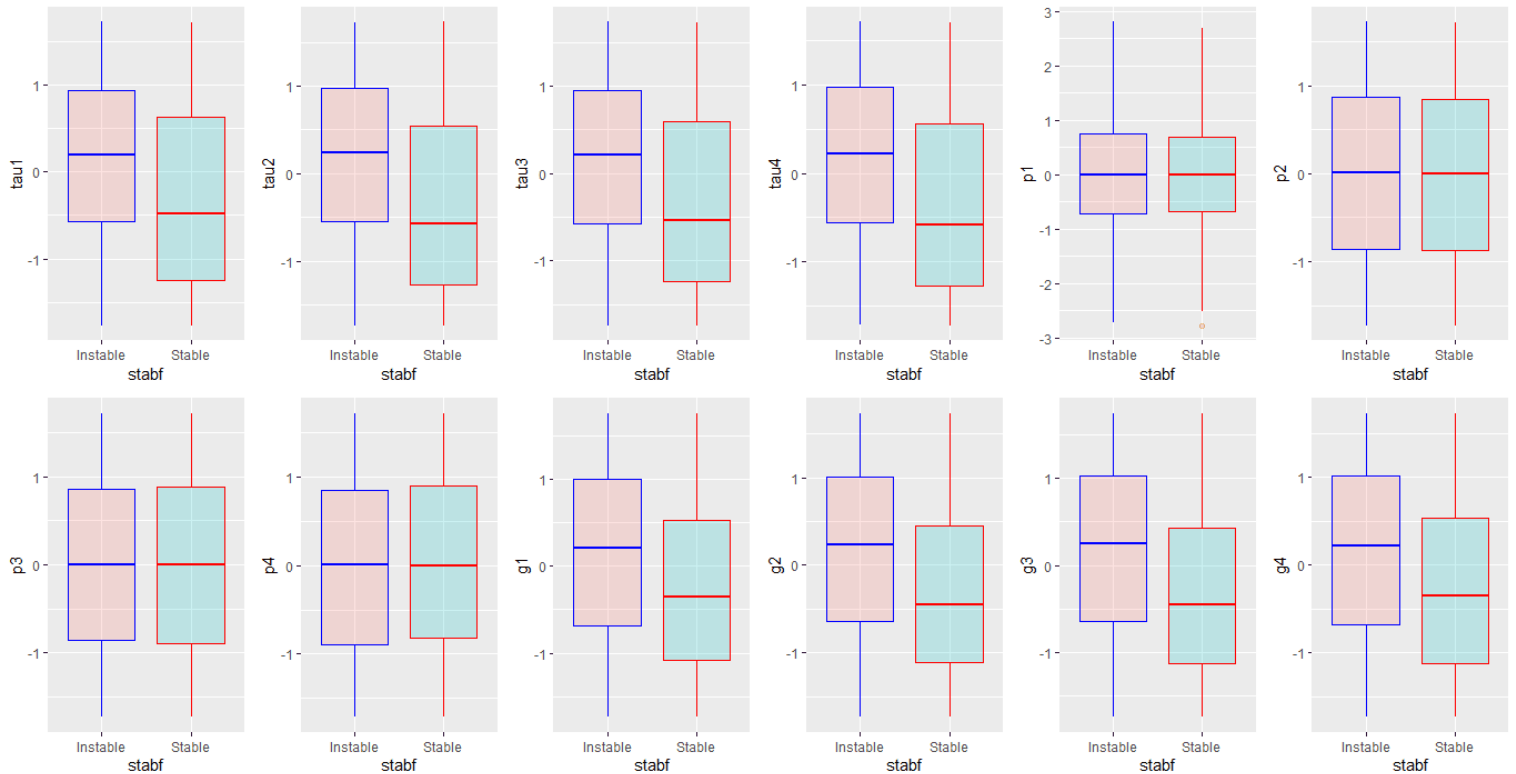


Dai Q-Q plots si nota che le variabili hanno delle code differenti da quelle delle distribuzioni teoriche. Tuttavia per verificare da un punto di vista statistico le nostre assunzioni, applichiamo lo Shapiro Test.

SHAPIRO TEST	tau 1	tau2	tau3	tau4	p2	p3	p4	g1	g2	g3	g4
Stable	0	0	0	0	0	0	0	0	0	0	0
Unstable	0	0	0	0	0	0	0	0	0	0	0

Come notiamo, i p-value, relativi all'ipotesi nulla ( $H_0$ : la distribuzione è normale) dello shapiro test confermano che le nostre esplicative non seguono un andamento normale.

Attraverso l'utilizzo dei boxplot condizionati notiamo che, generalmente, in corrispondenza di valori di esplicative più alti si è in presenza di reti elettriche instabili.



## METODI

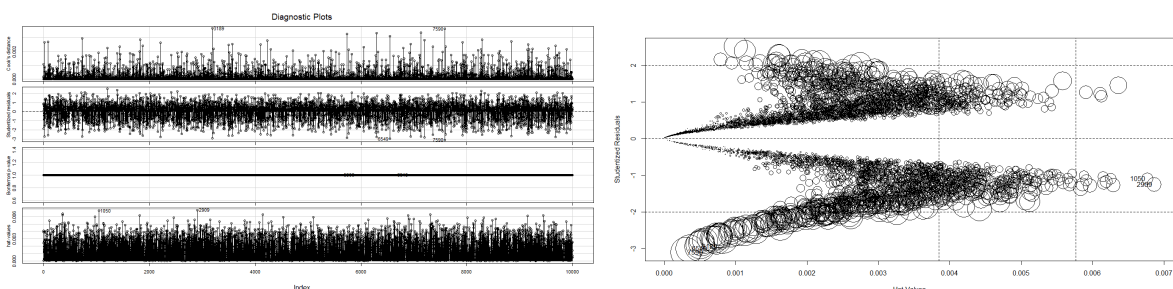
Avendo verificato in fase di pre processing che le variabili esplicative non si distribuiscono normalmente abbiamo escluso l'utilizzo di LDA e QDA come metodi di classificazione perché le assunzioni alla base di questi metodi non sono verificate empiricamente.

Procediamo, quindi, con altri metodi di classificazione.

Grazie a questi saremo in grado di classificare, tramite i valori di input, la stabilità di una rete elettrica.

## REGRESSIONE LOGISTICA

Come primo modello abbiamo deciso di testare il modello logistico utilizzando tutte le variabili rimanenti, avendo escluso p1 e stab precedentemente.



I grafici riguardanti i punti influenti o outliers sono di difficile interpretazione, data l'ampiezza campionaria dei dati. Tuttavia si nota che non ci sono dati estremi come avevamo già verificato durante le fasi del pre-processing.

Abbiamo deciso di utilizzare una selezione stepwise basata sull'AIC in modo da mantenere inalterata l'interpretazione delle variabili rimanenti. Questo processo ci porta ad eliminare le variabili p2, p3 e p4 in quanto poco significative.

Le rimanenti variabili risultano essere tutte altamente significative e la devianza risulta accettabile in riferimento ai gradi di libertà.

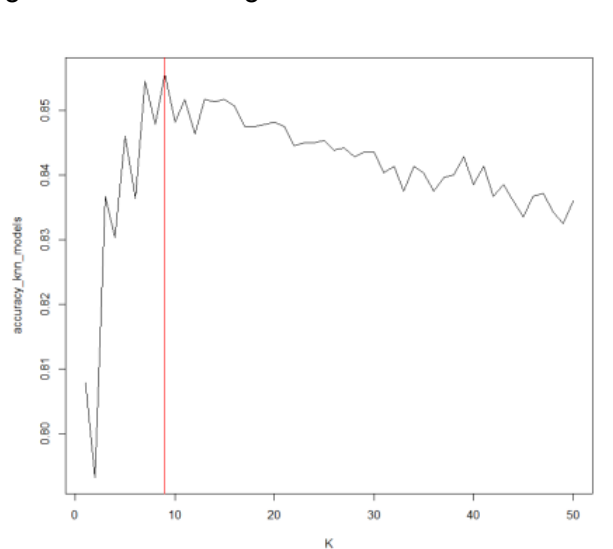
	intercept	t1	t2	t3	t4	g1	g2	g3	g4
coeff	1.07089	0.85976	0.90566	0.89915	0.92923	0.75579	0.79679	0.87957	0.77199
pval	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16

TEST DEVIANZA	DEVIANZA	GDL	PVAL
H0: modello adatto	6215.5	7991	1

### ALGORITMO K-NN

Come modello di classificazione alternativo abbiamo utilizzato l'algoritmo KNN, il quale (non deve essere applicato sul training set dato che in questo caso si otterrebbe un problema di overfitting con  $K=1$ , quindi) verrà applicato direttamente sul validation set sfruttando le osservazioni del training set come base.

Analizziamo l'accuracy del modello per ogni k compreso tra 1 e 50, riportata anche graficamente in seguito.



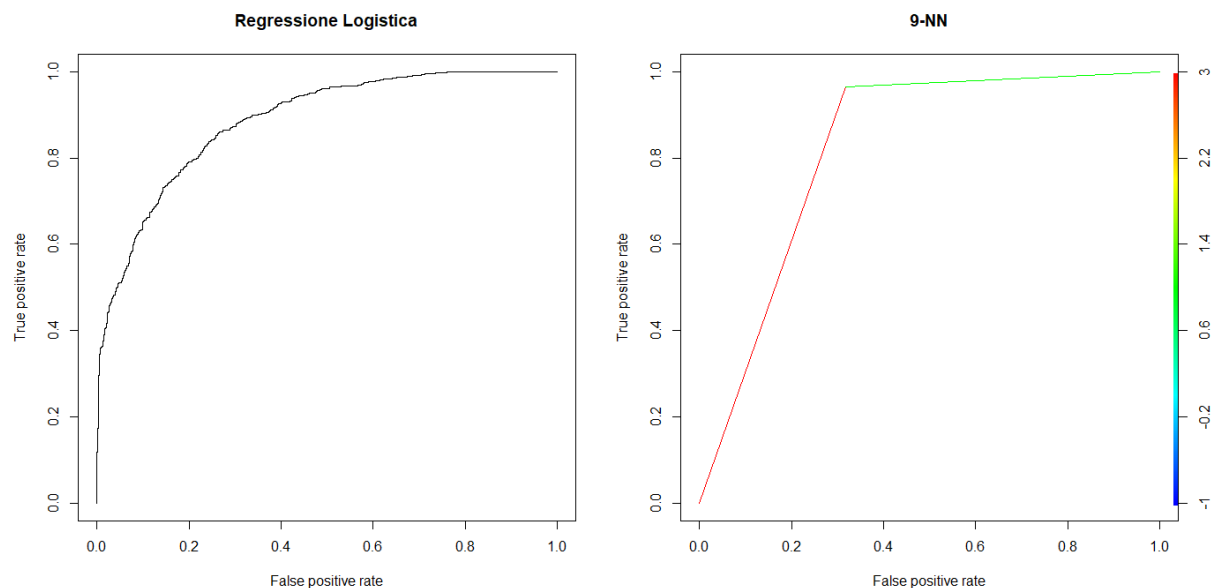
Come si denota dal grafico, il K che massimizza l'accuratezza del modello è uguale a 9.

## RISULTATI

Studiando la tabella dei coefficienti durante la regressione logistica, notiamo come, per valutare realmente il peso che hanno le covariate all'interno del modello, dobbiamo trovare gli Odds Ratio per ogni esplicativa (ad esempio per la variabile t1 l'O.R. è uguale a 2.362594). Nell'analisi otteniamo un'associazione positiva tra tutte le covariate e la variabile risposta (stabf).

RIS CONF MATRIX	REG LOG	K-NN
ACCURACY	0.8115	0.8625
SENSITIVITY	0.7566	0.9188
SPECIFICITY	0.8402	0.8416

Dalla tabella inserita, riusciamo a notare come l'Error Rate definito dal metodo K-NN (0.1375) è inferiore e di conseguenza migliore dello stesso indice definito invece dalla regressione logistica (0.1885). Questo risultato è dovuto essenzialmente al miglioramento dell'indice sensitivity che viene ottimizzato passando dal metodo parametrico a quello non-parametrico. Per quanto riguarda invece l'indice specificity, true negative rate, non si nota un miglioramento significativo.





## CONCLUSIONE

Le fonti di energia rinnovabili rappresentano la possibilità di un futuro con emissioni di CO2 significativamente ridotte. Tuttavia, a causa della loro flessibilità e della loro parcellizzazione, mettono a dura prova il sistema di reti elettriche.

Utilizzare un sistema centralizzato per la distribuzione dell'energia richiederebbe, oltre ad una rete elettrica più sofisticata, un'elaborazione di dati ed un lavoro computazionale elevatissimo. Un sistema decentralizzato e avanzato dal punto di vista del machine learning, invece, permetterebbe di avere reti elettriche stabili e al contempo di stabilire il prezzo dell'energia e comunicarlo direttamente ai consumatori, senza collezionare e processare un numero troppo elevato di dati.

## REFERENZE

- New Journal of Physics, Volume 17, Issue 1, article id. 015002 (2015).
- Energyup.tech, Smart grid: cos'è e come funziona la rete elettrica intelligente in Italia, 15 Giu 2020, Gianluigi Torchiani
- Arzamasov, Vadim, Klemens Bahm, and Patrick Jochem. 'Towards Concise Models of Grid Stability.' Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), 2018