

UNIVERSITA' DEGLI STUDI DI BARI «ALDO MORO»

DIPARTIMENTO DI INFORMATICA

Corso di laurea in «informatica e tecnologie per la produzione del software»



Tesi di laurea

PREDIZIONE DELLA RADIAZIONE SOLARE ATTRAVERSO ALGORITMI DI MACHINE LEARNING

Relatore: Donato Impedovo

Laureando: Simone Sorrenti

ANNO ACCADEMICO: 2017-2018

INTRODUZIONE

- Predizione dell'indice della radiazione solare per diversi orizzonti temporali:
 1. Orario
 2. Giornaliero
 3. Mensile
- Equazione di Penman-Monteith per determinare l'evapotraspirazione potenziale:

$$ET_0 = \frac{0,408 \Delta (R_n - G) + \gamma \left(\frac{900}{T_k} \right) U_2 (e_a - e_d)}{\Delta + \gamma (1 + 0,34 U_2)}$$

Dove:

ET_0 = evapotraspirazione di riferimento (mm d^{-1})

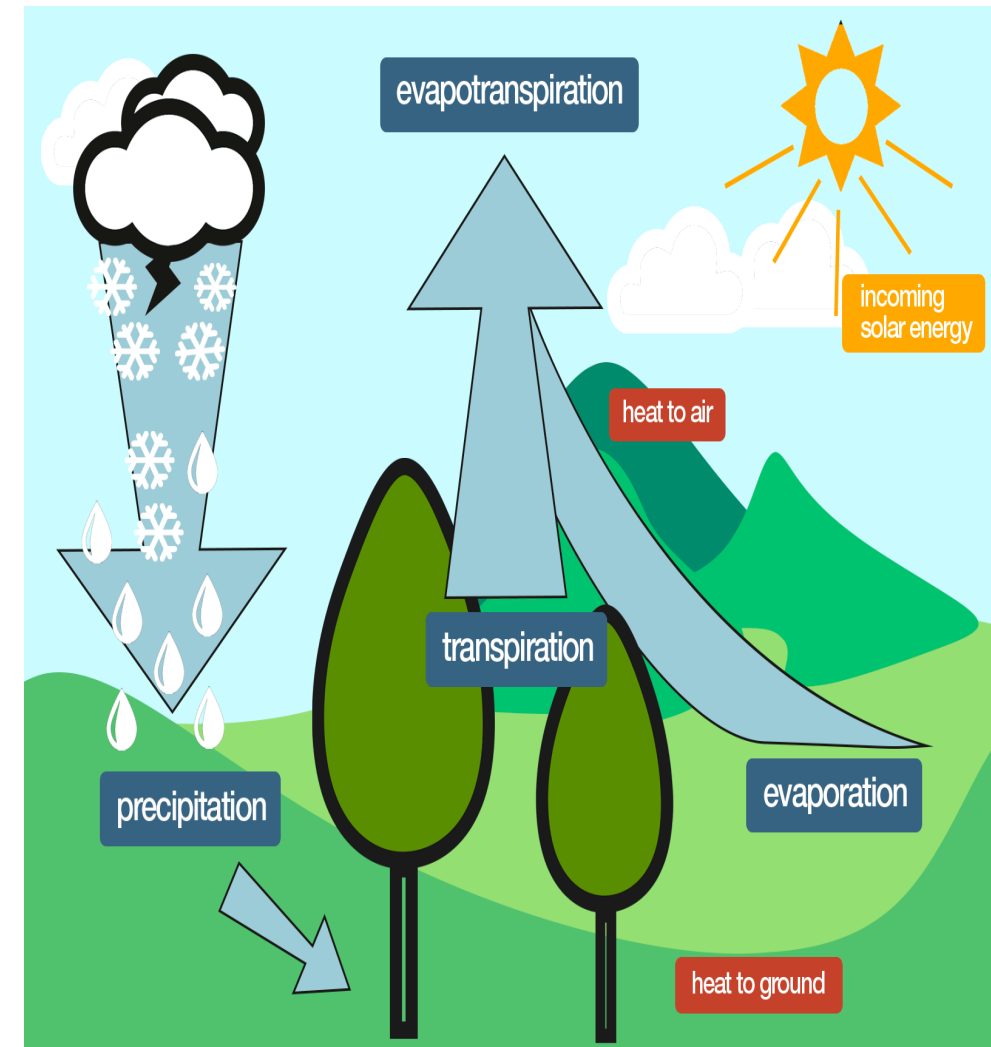
Δ = tensione di vapore saturo in funzione della temperatura ($\text{kPa } ^\circ\text{C}^{-1}$)

R_n = radiazione netta ($\text{MJ m}^{-2} \text{d}^{-1}$)

T_k = temperatura assoluta media a 2 m dal suolo ($^\circ\text{K}$)

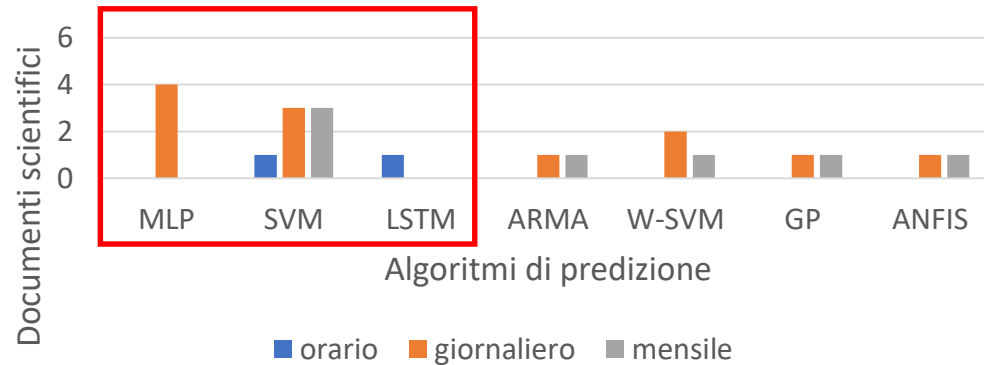
U_2 = velocità del vento a 2 m dal suolo (m s^{-1})

G = flusso di calore dal suolo ($\text{MJ m}^{-2} \text{d}^{-1}$)

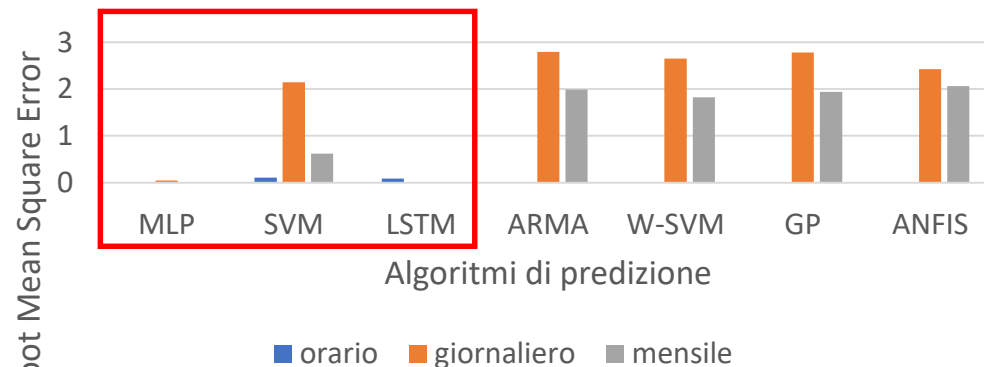


STATO DELL'ARTE

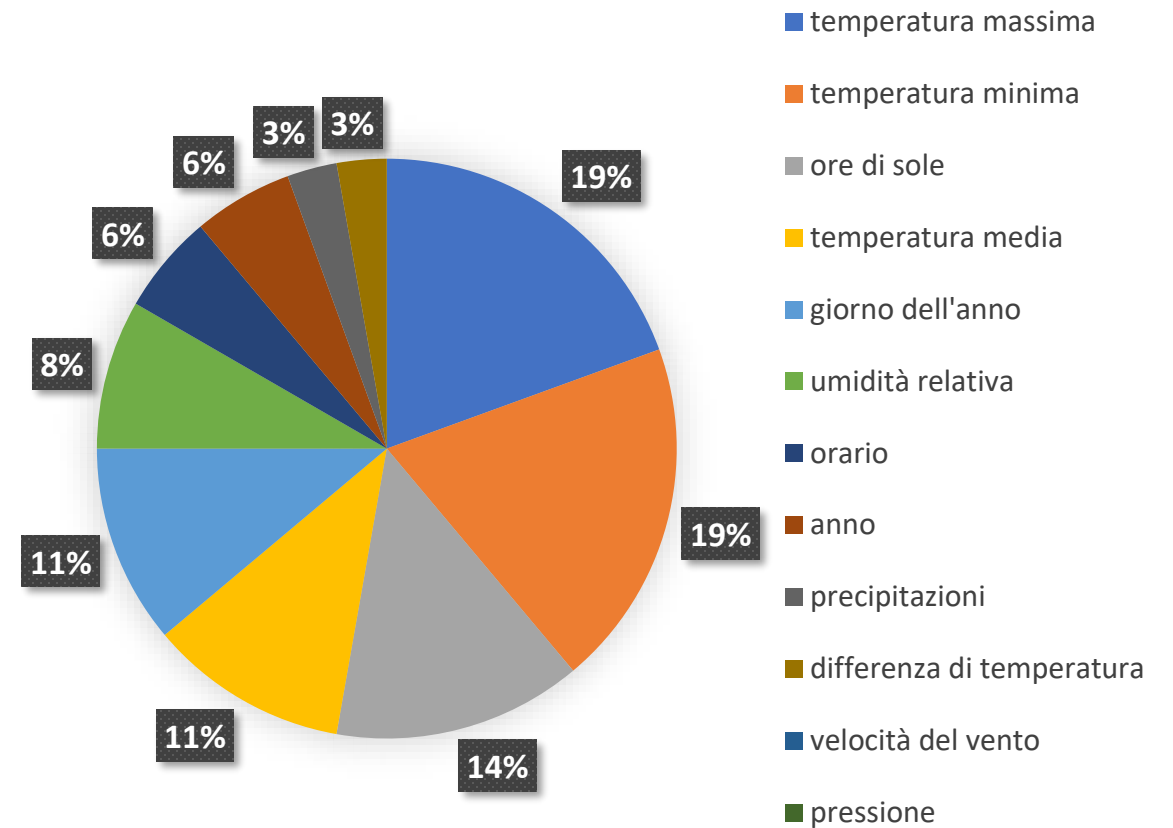
Algoritmi di predizione per la radiazione solare maggiormente utilizzati



Accuratezze delle predizioni della radiazione solare



Frequenza d'uso dei parametri meteorologici per la predizione della radiazione solare



METODI E METRICHE

- Per addestrare i modelli predittivi sono stati forniti in input:
 - ✓ minimo un parametro
 - ✓ massimo 4 parametridei primi 4 parametri più rilevanti secondo gli algoritmi di Feature Selection

- Algoritmi di Feature Selection utilizzati



Pearson
Correlation

Recursive Feature
Elimination con
SVM

Random Forest

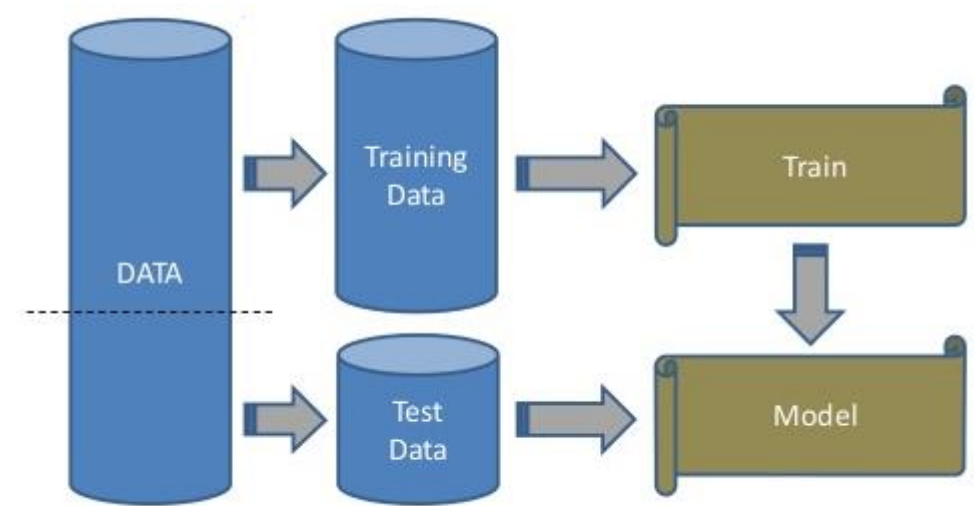
- Metrica di errore usata



$$RMSE(Root\ Mean\ Square\ Error) = \sqrt{\frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N}}$$

DATA SET

- Data set fornito dall'azienda SysMan:
 - ✓ Rilevazioni dei parametri meteorologici: orarie
 - ✓ Luogo: stazione meteorologica di Molfetta (id=186)
 - ✓ Periodo: 1 luglio 2017 - 28 Febbraio 2018
- Data set suddiviso in Training set e Testing set:
 1. ~70% Training set e ~30% Testing set
 2. ~80% Training set e ~20% Testing set



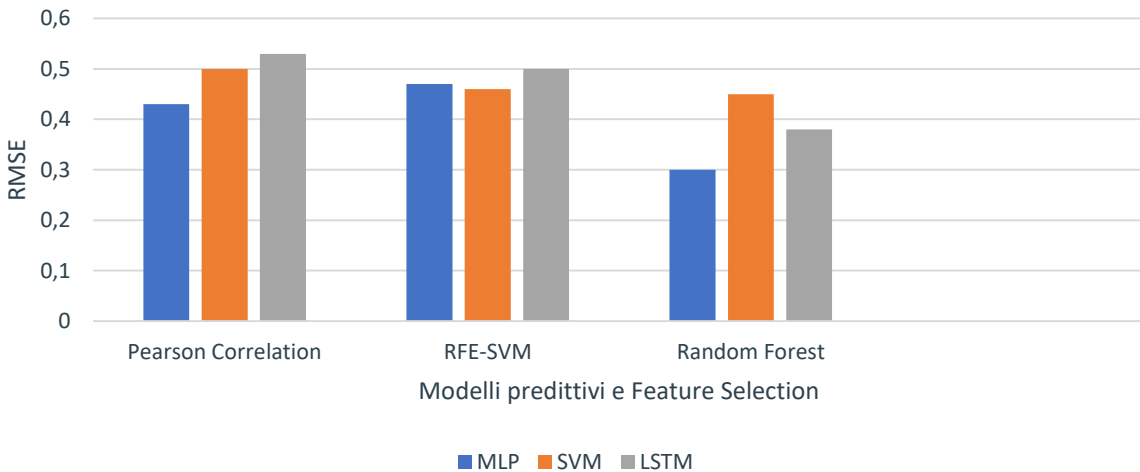
Orizzonte	~70% Training set	~80% Training set	~30% Testing set	~20% Testing set
Orario	1 Luglio 2017 – 17 Dicembre 2017	1 Luglio 2017 – 10 Gennaio 2018	18 Dicembre 2017 – 28 Febbraio 2018	11 Gennaio 2018 – 28 Febbraio 2018
Giornaliero				
Mensile	Luglio 2017-Dicembre 2017		Gennaio 2018 – Febbraio 2018	

RISULTATI SPERIMENTALI: ORIZZONTE ORARIO

RANKING	PEARSON CORRELATION	RECURSIVE FEATURE ELIMINATION CON SVM	RANDOM FOREST
1	temperatura minima	velocità del vento	temperatura minima
2	temperatura media	ore di sole	orario
3	temperatura massima	temperatura massima	velocità del vento
4	umidità	temperatura media	ore di sole
5	velocità del vento	differenza di temperatura	giorno dell'anno
6	ore di sole	temperatura minima	pressione
7	orario	precipitazioni	umidità
8	giorno dell'anno	anno	temperatura massima
9	anno	orario	differenza di temperatura
10	differenza di temperatura	umidità	temperatura media
11	precipitazioni	pressione	precipitazioni
12	pressione	giorno dell'anno	anno

ACCURATEZZA DELLE PREDIZIONI (RMSE)				
FEATURE SELECTION	PARAMETRI IN INPUT	MLP	SVM	LSTM
Pearson Correlation	Tmin	0.456	0.522	0.557
	Tmin + Tmed	0.446	0.518	0.574
	Tmin + Tmed + Tmax	0.441	0.517	0.574
	Tmin + Tmed + Tmax + RH_med	0.392	0.475	0.432
Recursive feature elimination (SVM)	WS	0.575	0.44	0.437
	WS + ore_sole	0.459	0.449	0.450
	WS + ore_sole + Tmax	0.426	0.486	0.564
	WS + ore_sole + Tmax + Tmed	0.431	0.489	0.570
Random Forest	Tmin	0.456	0.522	0.557
	Tmin + date_time_hour	0.271	0.508	0.349
	Tmin + date_time_hour + WS	0.265	0.488	0.335
	Tmin + date_time_hour + WS + ore_sole	0.243	0.291	0.313

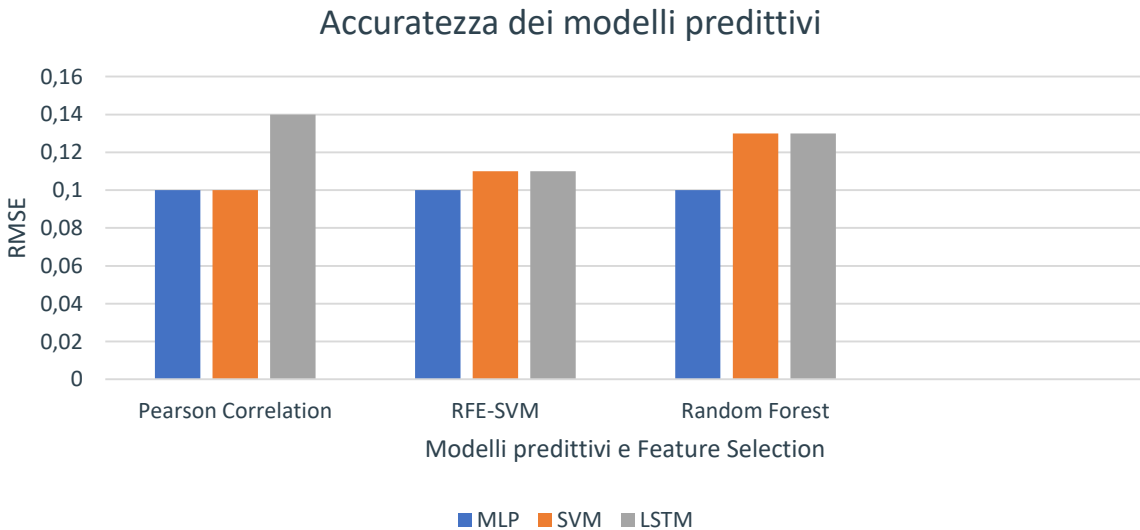
Accuratezza dei modelli predittivi



- Modello predittivo più accurato: Multi-Layer Perceptron
- Algoritmo di Feature Selection migliore: Random Forest
- I parametri meteorologici che permettono di avere predizioni orarie più accurate sono: la temperatura minima, l'orario della rilevazione, la velocità del vento e le ore di sole

RISULTATI SPERIMENTALI: ORIZZONTE GIORNALIERO

RANKING	PEARSON CORRELATION	RECURSIVE FEATURE ELIMINATION CON SVM	RANDOM FOREST
1	ore di sole	ore di sole	ore di sole
2	temperatura minima	precipitazioni	temperatura massima
3	temperatura media	velocità del vento	giorno dell'anno
4	temperatura massima	temperatura minima	umidità
5	umidità	differenza di temperatura	differenza di temperatura
6	differenza di temperatura	temperatura massima	temperatura minima
7	giorno dell'anno	anno	precipitazioni
8	anno	umidità	pressione
9	precipitazioni	pressione	temperatura media
10	velocità del vento	temperatura media	velocità del vento
11	pressione	giorno dell'anno	anno



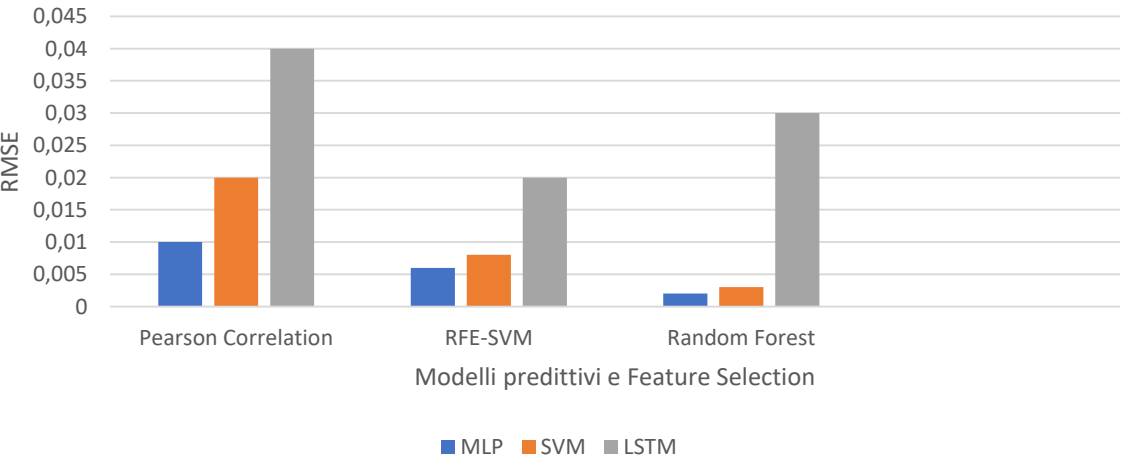
ACCURATEZZA DELLE PREDIZIONI (RMSE)				
FEATURE SELECTION	PARAMETRI IN INPUT	MLP	SVM	LSTM
Pearson Correlation	ore_sole	0.127	0.131	0.100
	ore_sole + Tmin	0.103	0.103	0.168
	ore_sole + Tmin + Tmed	0.104	0.0991	0.161
	ore_sole + Tmin + Tmed + Tmax	0.0941	0.0959	0.170
Recursive feature elimination (SVM)	ore_sole	0.127	0.131	0.100
	ore_sole + rain	0.101	0.107	0.102
	ore_sole + rain + WS	0.100	0.105	0.107
	ore_sole + rain + WS + Tmin	0.0956	0.102	0.134
Random Forest	ore_sole	0.127	0.131	0.100
	ore_sole + Tmax	0.103	0.101	0.170
	ore_sole + Tmax + giorno dell'anno	0.110	0.153	0.147
	ore_sole + Tmax + giorno dell'anno+ RH_med	0.0928	0.155	0.117

- Modello predittivo più accurato: Multi-Layer Perceptron
- Algoritmo di Feature Selection migliore: Random Forest
- I parametri meteorologici che permettono di avere predizioni giornaliere più accurate sono: le ore di sole, la temperatura massima, il giorno dell'anno e l'umidità relativa

RISULTATI SPERIMENTALI: ORIZZONTE MENSILE

RANKING	PEARSON CORRELATION	RECURSIVE FEATURE ELIMINATION CON SVM	RANDOM FOREST
1	ore di sole	umidità	temperatura minima
2	differenza di temperatura	temperatura massima	temperatura massima
3	temperatura massima	temperatura media	umidità
4	temperatura media	temperatura minima	differenza di temperatura
5	temperatura minima	ore di sole	pressione
6	umidità	pressione	temperatura media
7	precipitazioni	differenza di temperatura	precipitazioni
8	velocità del vento	precipitazioni	ore di sole
9	pressione	velocità del vento	velocità del vento
10	anno	anno	anno

Accuratezza dei modelli predittivi



ACCURATEZZA DELLE PREDIZIONI (RMSE)				
FEATURE SELECTION	PARAMETRI IN INPUT	MLP	SVM	LSTM
FEATURE SELECTION Pearson Correlation	ore_sole	0.0655	0.0537	0.0348
	ore_sole + DiffTemp	0.01	0.0532	0.0506
	ore_sole + DiffTemp + Tmax	0.0002	0.0002	0.0348
	ore_sole + DiffTemp + Tmax + Tmin	0.000808	0.00026	0.0408
Recursive feature elimination (SVM)	RH_med	0.0224	0.0203	0.0255
	RH_med + Tmax	0.000899	0.00758	0.0266
	RH_med + Tmax + Tmed	0.000874	0.00479	0.0273
	RH_med + Tmax + Tmed + Tmin	0.000441	0.00287	0.00832
Random forest	Tmin	0.00304	0.00245	0.0434
	Tmin + Tmax	0.00724	0.00197	0.0511
	Tmin + Tmax + RH_med	0.000469	0.00444	0.0270
	Tmin + Tmax + RH_med + DiffTemp	0.000929	0.00474	0.0184

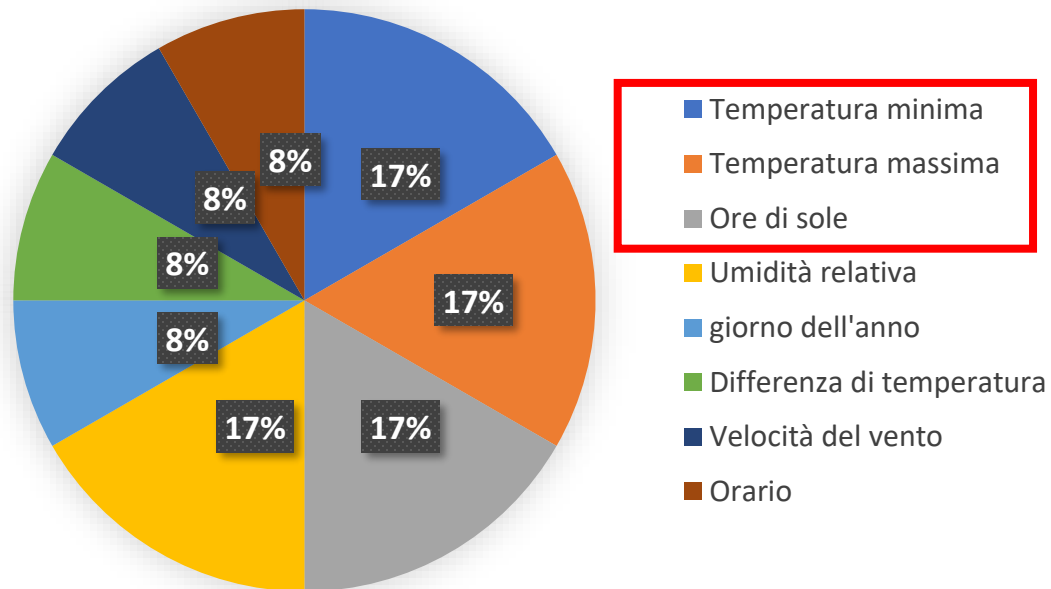
- Modello predittivo più accurato: Multi-Layer Perceptron
- Algoritmo di Feature Selection migliore: Random Forest
- I parametri meteorologici che permettono di avere predizioni giornaliere più accurate sono: la temperatura minima, la temperatura massima, l'umidità relativa e la differenza di temperatura

CONCLUSIONI

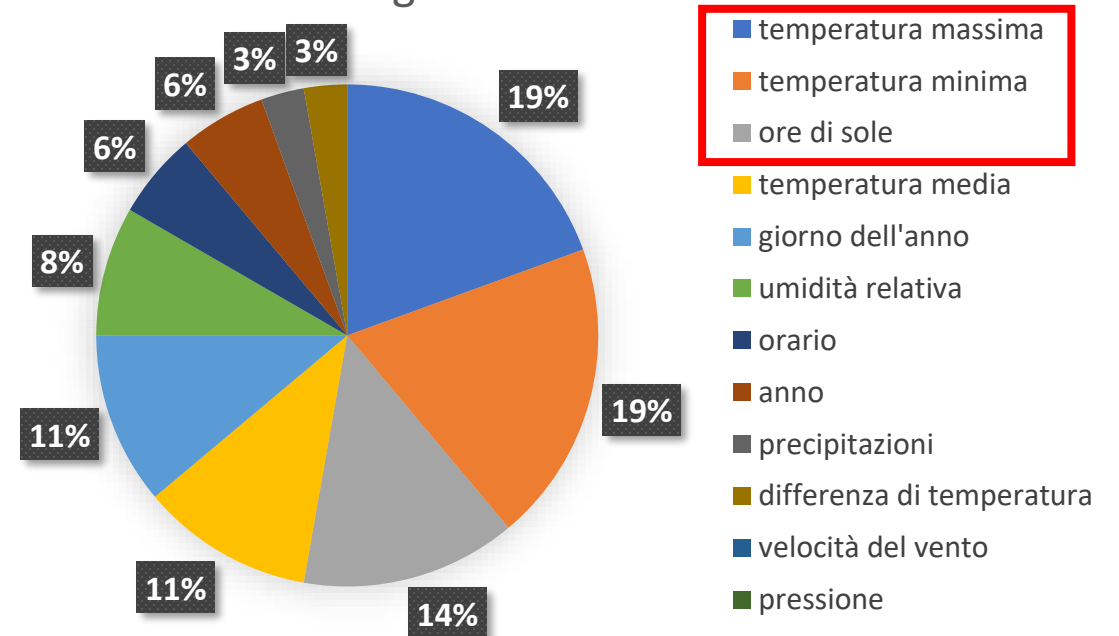
Indipendentemente dall'orizzonte temporale:

- Il modello predittivo più accurato: Multi-Layer Perceptron
- L'algoritmo di Feature Selection migliore: Random Forest
- I parametri meteorologici che permettono di avere predizioni più accurate sono: la temperatura minima, la temperatura massima e le ore di sole

Frequenza d'uso dei parametri meteorologici delle predizioni più accurate



Frequenza d'uso dei parametri meteorologici delle fonti scientifiche



SVILUPPI FUTURI

Un possibile futuro sviluppo dello studio potrebbe essere quello di:

- Predire la radiazione solare per le stazioni meteorologiche adiacenti
- Predire la temperatura, la velocità del vento e l'umidità relativa
- Determinare l'evapotraspirazione, mediante l'equazione di Penman-Monteith, utilizzando i valori predetti dei parametri meteorologici richiesti

GRAZIE A TUTTI PER L'ATTENZIONE