Stellar Classification

Machine Learning - 17/02/2020

Christian Bernasconi 816423
Riccardo Pozzi 807857
Marco Ripamonti 806785

Dataset

Il dataset scelto comprende un insieme di 240 istanze, ciascuna di esse descritta da 7 diverse features.

| Temp	Lum	Rad	AbsMagn	Color (19 --> 11 levels)	SpectrClass (7 levels)	Type (6 levels)
3068	0.002400	0.1700	16.12	red	M	Supergiant
3042	0.000500	0.1542	16.60	white	O	WhiteDwarf
2600	0.000300	0.1020	18.70	blue white	F	Mainsequence
...	...	...	...	...	...	...

Analisi esplorativa

Distribuzione del target

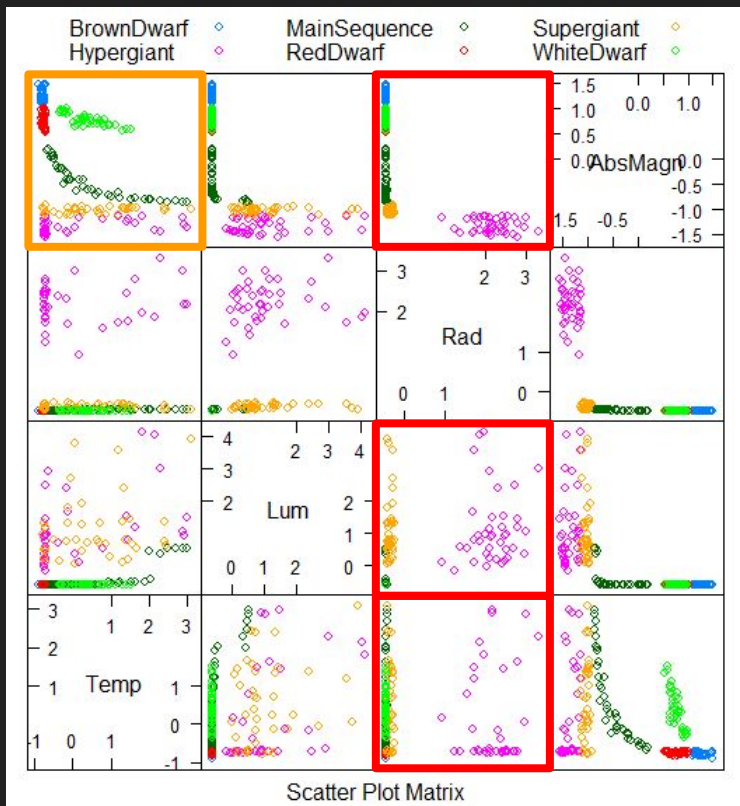| Type	Distribuzione
BrownDwarf	40
Hypergiant	40
Mainsequence	40
RedDwarf	40
Supergiant	40
WhiteDwarf	40
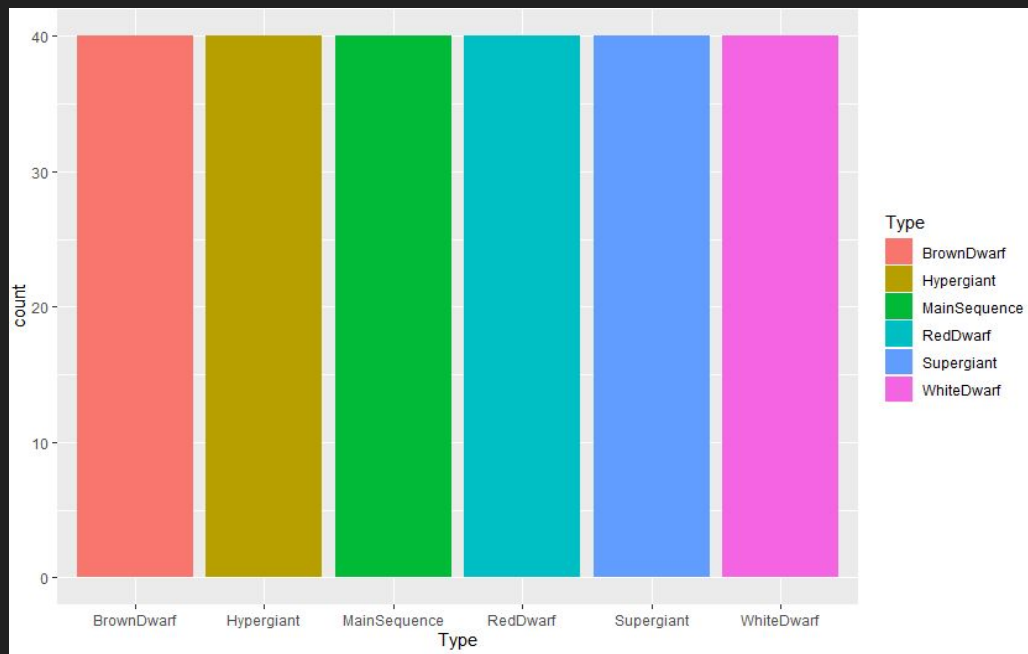
Correlazione features numeriche
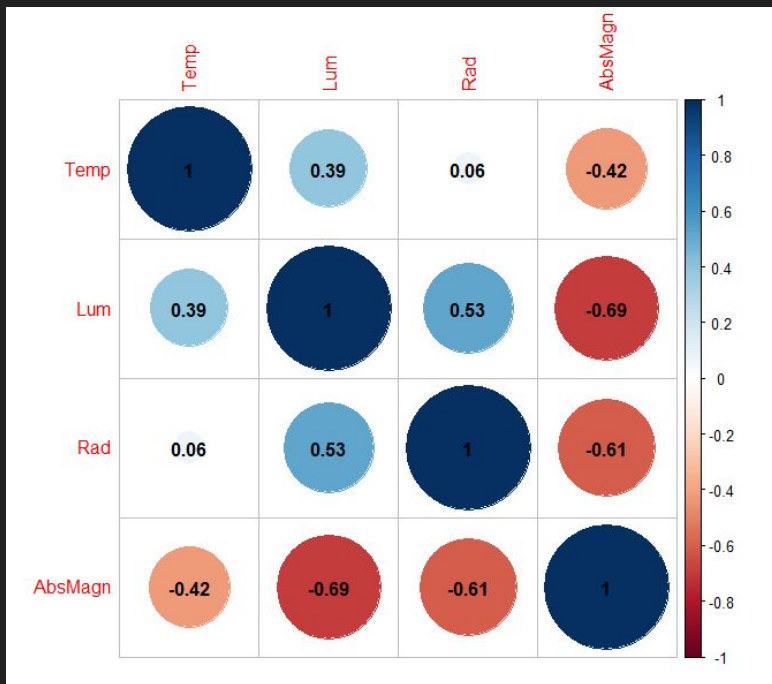
- no valori molto vicini a ±1
- poche features

riduzione dimensionalità non necessaria

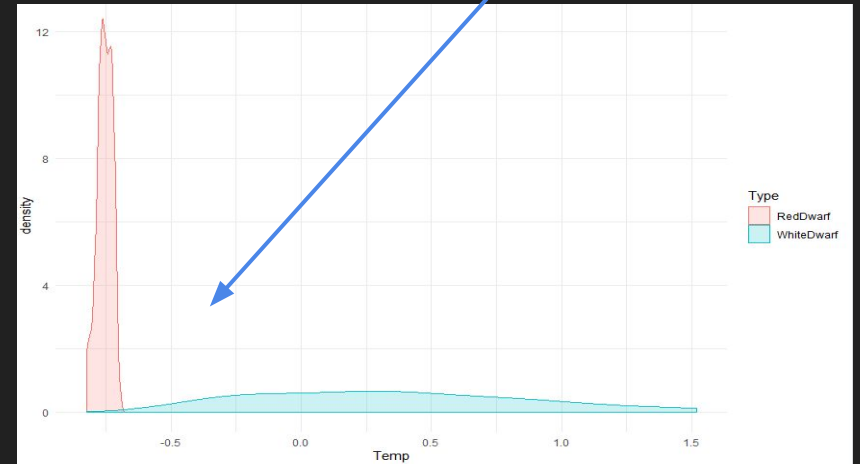# Verifica separabilità classi per coppie di features
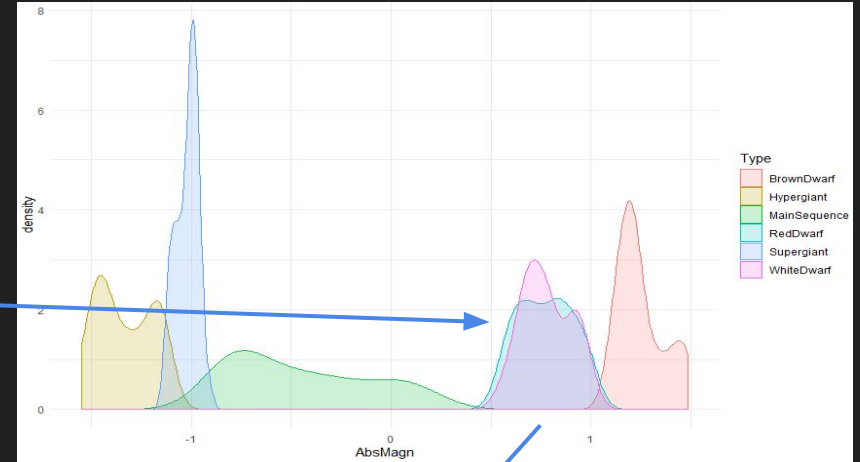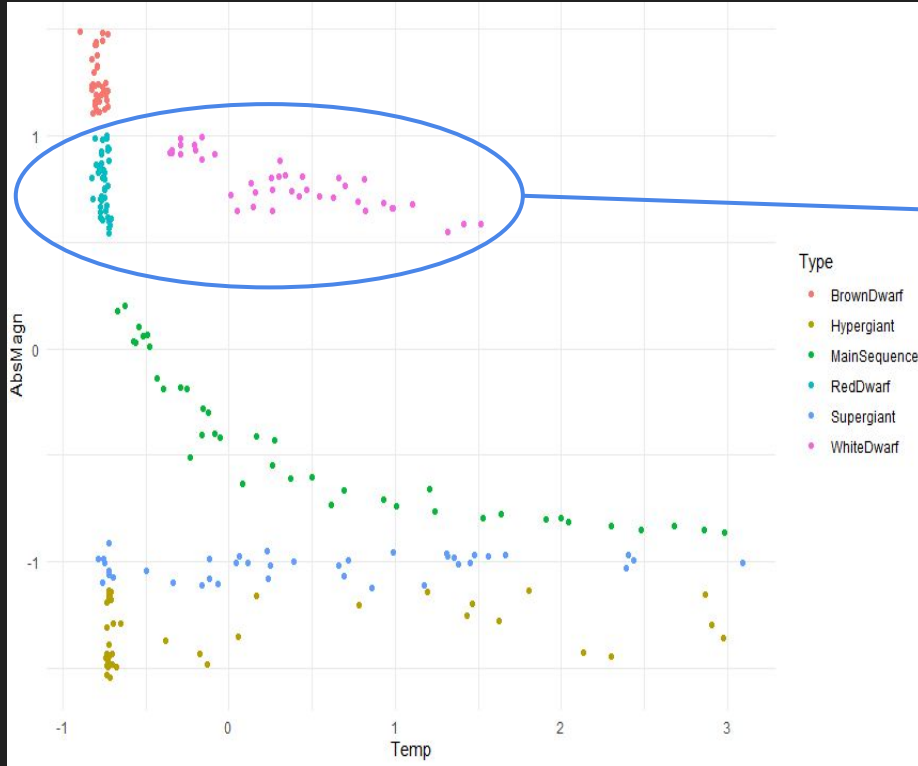

Scatter Plot Matrix

Osservazioni:

- AbsMagn + Temp separano abbastanza bene tutte le classi
- Rad identifica molto bene la classe Hypergiant in qualsiasi combinazione
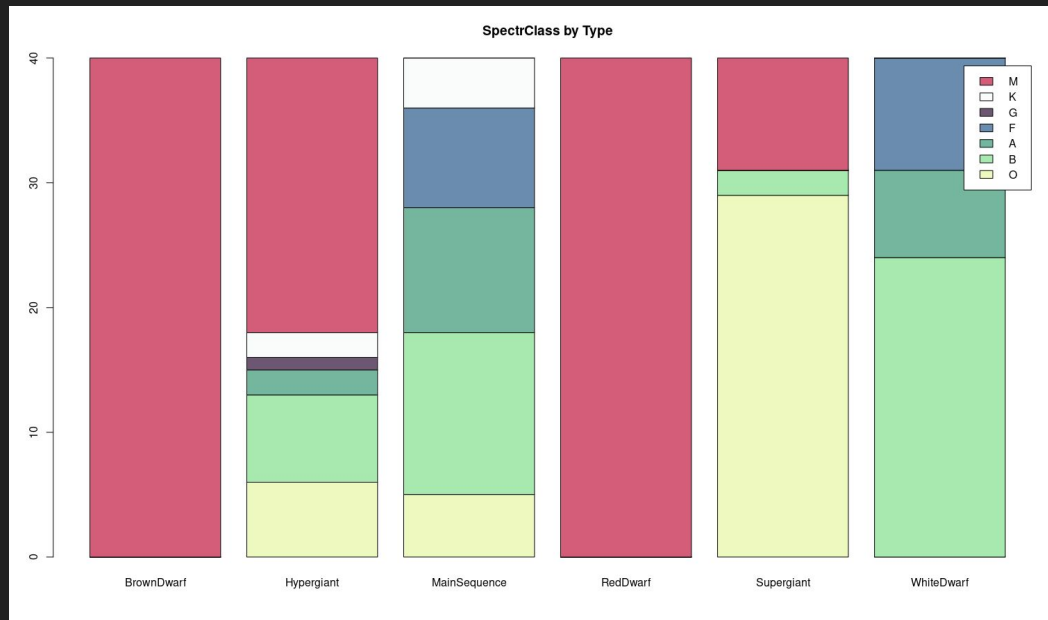
un algoritimo che classifica in base a iperpiani potrebbe lavorare bene

Dettaglio distribuzioni

Distribuzione target - SpectrClass

Osservazioni:

● da solo non fornisce molte informazioni
● come nel caso precedente, può essere usato in aiuto ad AbsMagn*per distinguere RedDwarf e WhiteDwarf

*Harvard Spectral Classification = AbsMagn + SpectrClass

Harvard Spectral Classification
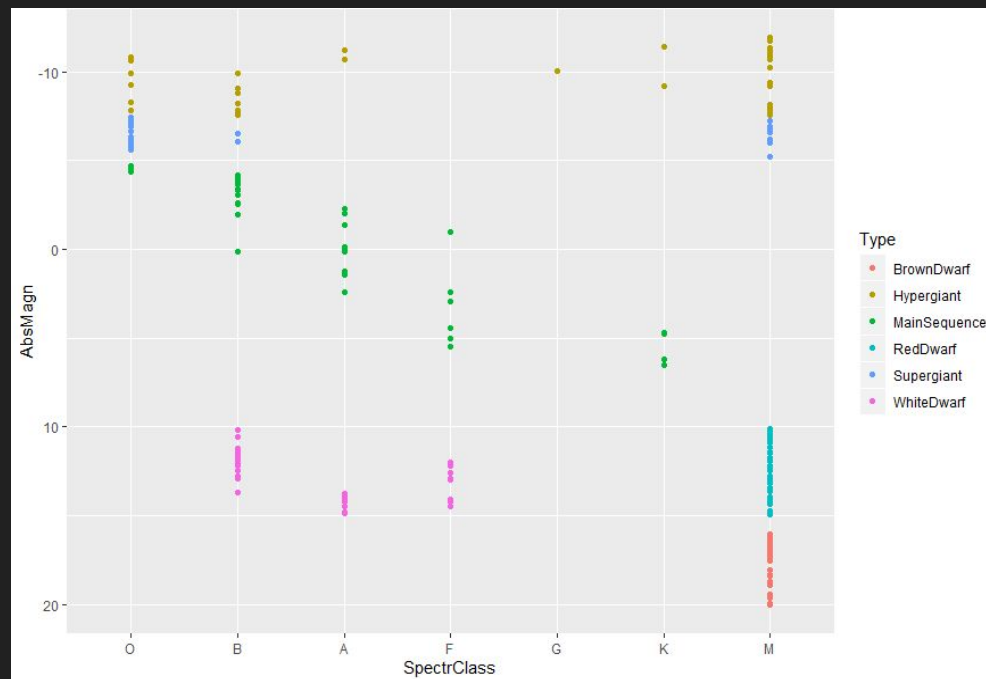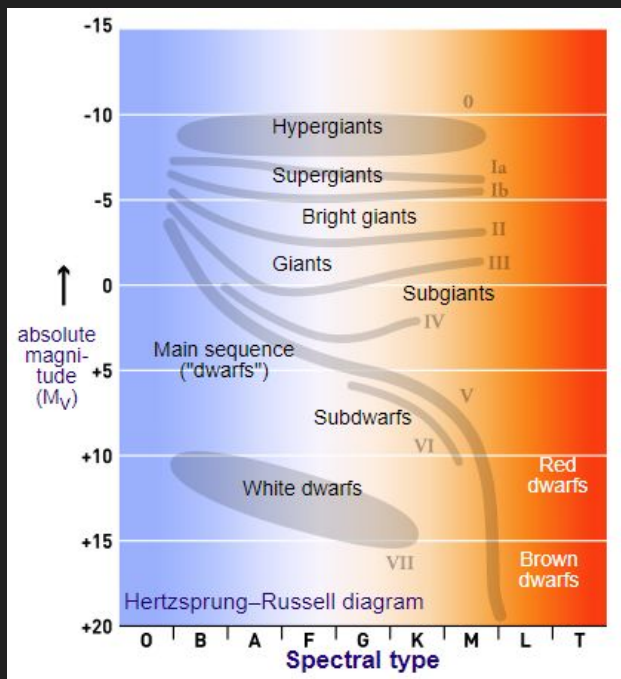
Correlazione features categoriche
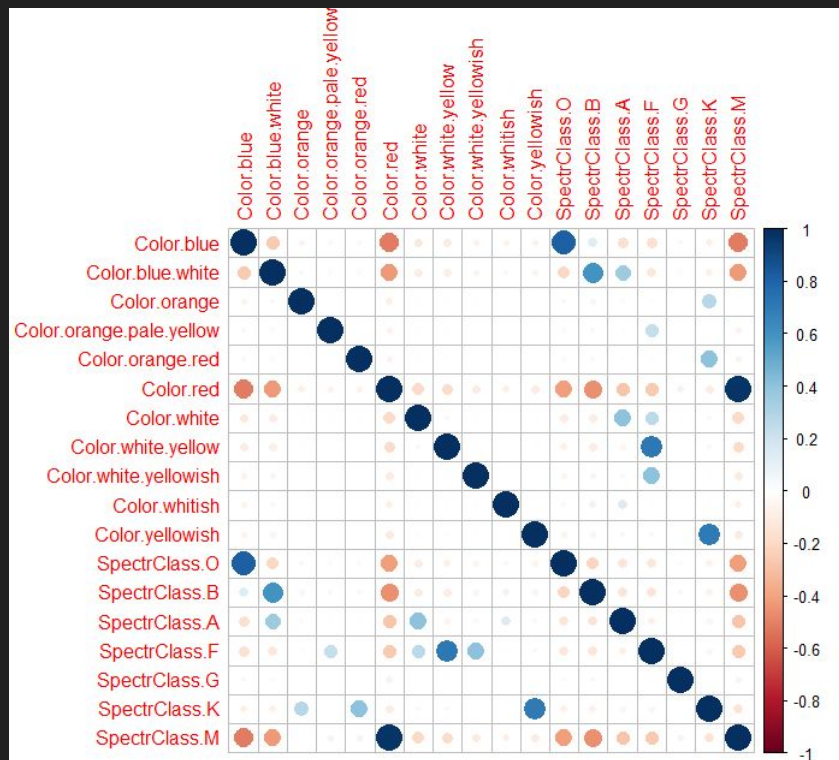
Osservazioni:

- presenza di coppie fortemente correlate: (O, blue), (B, blue white), (F, white yellow), (M, red), …
- alcuni colori inconsistenti (white/whitish, yellow/yellowish) portano a distribuire la correlazione su più spettri

Modelli adottati

Scelta modelli

- **SVM**

Scatter Plot Matrix

Scelta modelli

- SVM
- **Decision Tree**

Scelta modelli

- SVM
- Decision Tree
- **Random Forest**

Workflow

SVM

Approccio scartato:

Features: AbsMagn + Temp
Kernel: linear
Cost: 1; calcolato da 10-fold cv

Accuracy: 0.93

```
             Reference
Prediction    BrownDwarf Hypergiant MainSequence RedDwarf Supergiant whiteDwarf
  BrownDwarf      15          0           0          0         0          0
  Hypergiant       0         11           0          0         0          0
  MainSequence     0          0          10          0         1          0
  RedDwarf         0          0           0          9         0          0
  Supergiant       0          4           0          0        16          0
  whiteDwarf       0          0           0          0         0          9
```
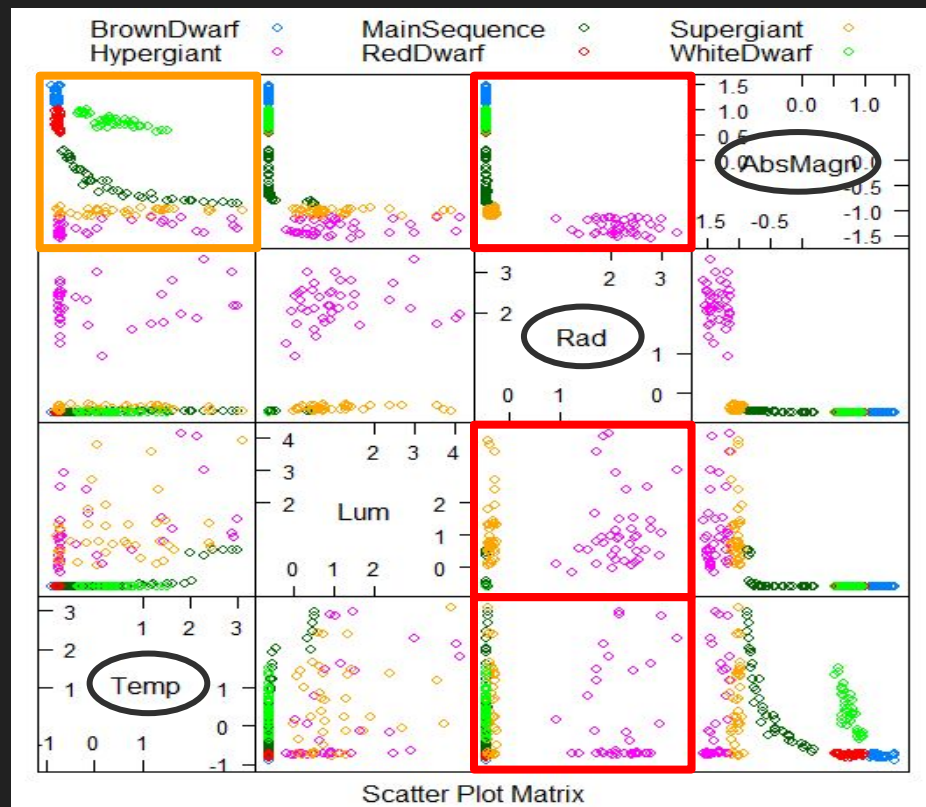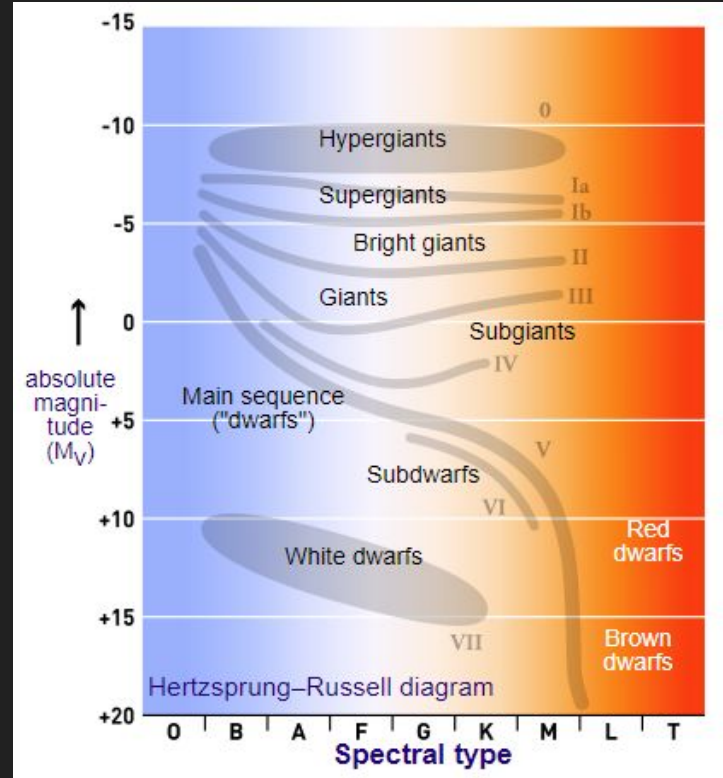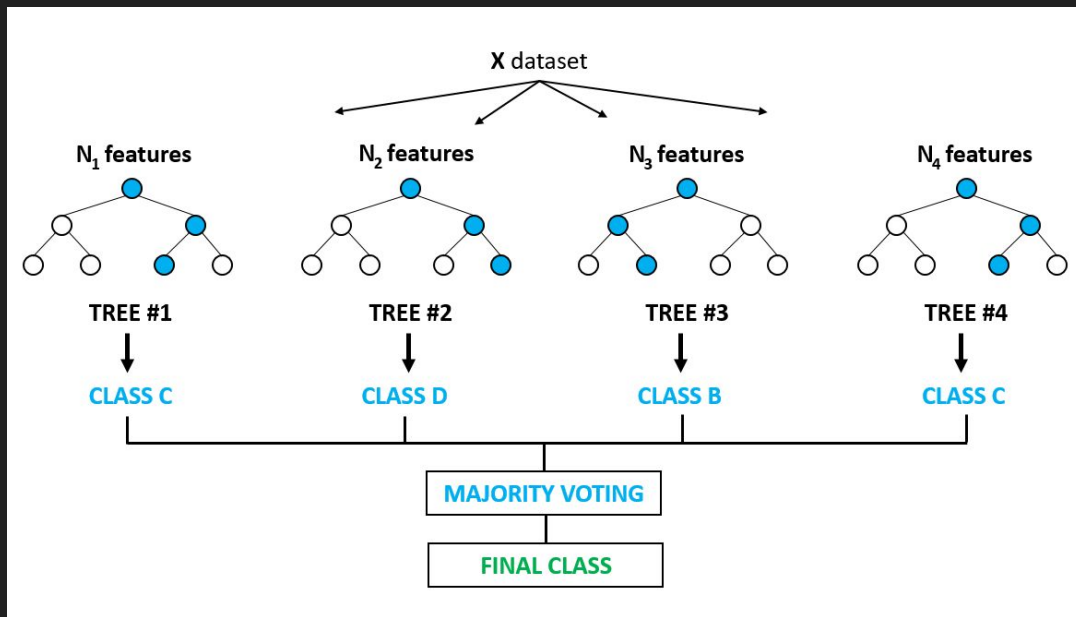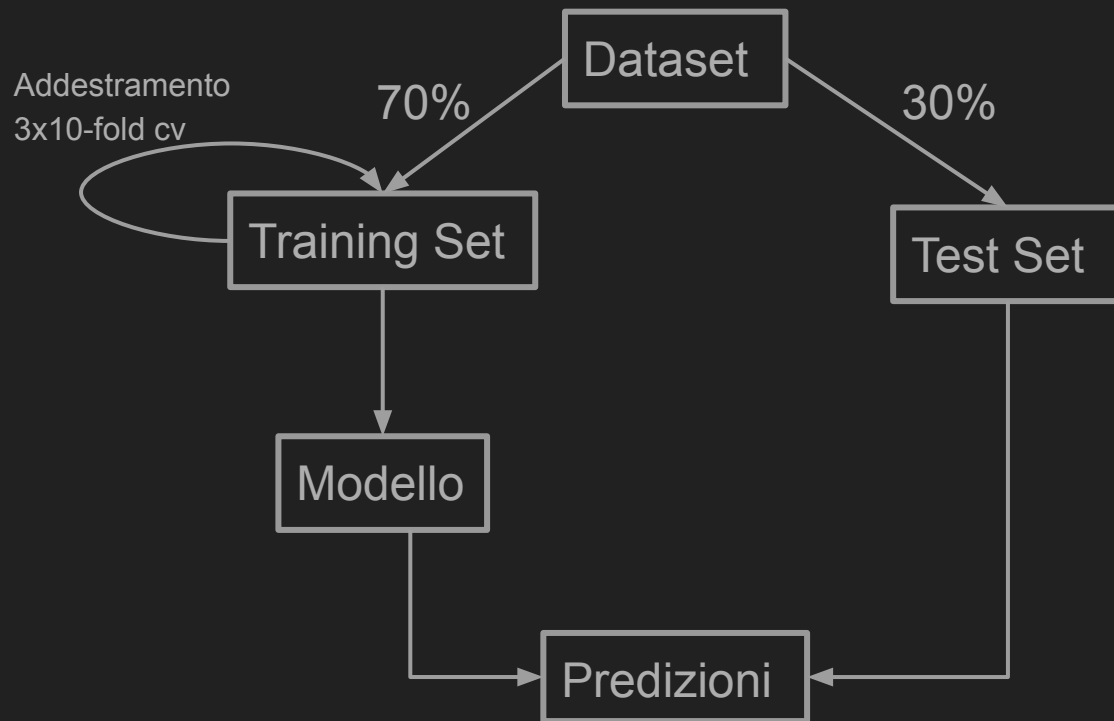
Approccio finale:

Features: AbsMagn + Temp + Rad
Kernel: linear
Cost: 1; calcolato da 10-fold cv

Accuracy: 0.98

```
             Reference
Prediction    BrownDwarf Hypergiant MainSequence RedDwarf Supergiant whiteDwarf
  BrownDwarf      15          0           0          0         0          0
  Hypergiant       0         15           0          0         0          0
  MainSequence     0          0          10          0         1          0
  RedDwarf         0          0           0          9         0          0
  Supergiant       0          0           0          0        16          0
  whiteDwarf       0          0           0          0         0          9
```

SVM - Risultati

	Class: BrownDwarf	Class: Hypergiant
Precision	1.0	1.0
Recall	1.0	1.0
F1	1.0	1.0
Prevalence	0.2	0.2
Detection Rate	0.2	0.2
Detection Prevalence	0.2	0.2
Balanced Accuracy	1.0	1.0
	Class: MainSequence	Class: RedDwarf
Precision	0.9091	1.00
Recall	1.0000	1.00
F1	0.9524	1.00
Prevalence	0.1333	0.12
Detection Rate	0.1333	0.12
Detection Prevalence	0.1467	0.12
Balanced Accuracy	0.9923	1.00
	Class: Supergiant	Class: WhiteDwarf
Precision	1.0000	1.00
Recall	0.9412	1.00
F1	0.9697	1.00
Prevalence	0.2267	0.12
Detection Rate	0.2133	0.12
Detection Prevalence	0.2133	0.12
Balanced Accuracy	0.9706	1.00

Legend: BrownDwarf, Hypergiant, MainSequence, RedDwarf, Supergiant, WhiteDwarf, Macro, Micro, SVM (axes: FPR, TPR)
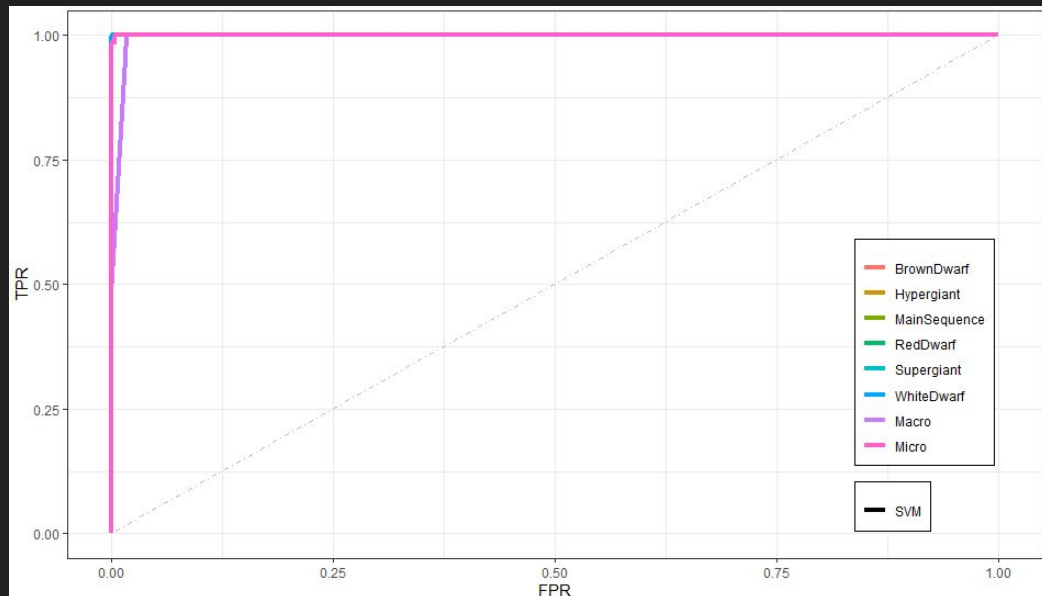
Macro:
precision: 0.9848485 - **recall:** 0.9901961 - **f-measure:** 0.9870130 - **AUC:** 0.9959891

Decision Tree

Prove varie:

Features: AbsMagn + SpectrClass +...
Type: CART
CP: 0.172; calcolato da 10-fold cv

Accuracy: ≤ 0.73

Approccio finale:

Features: AbsMagn + SpectrClass
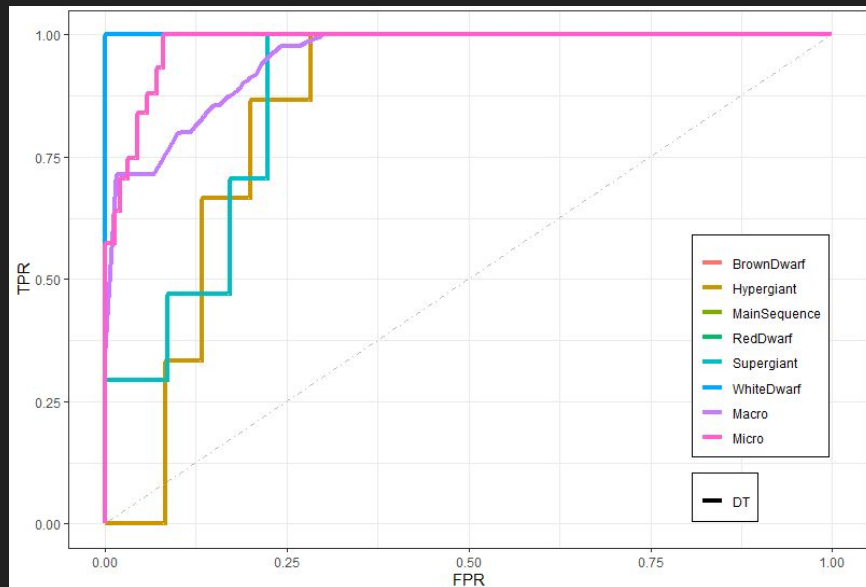**Type:** CART
CP: 0.172; calcolato da 10-fold cv

Accuracy: 0.73

```
                 Reference
Prediction    BrownDwarf Hypergiant MainSequence RedDwarf Supergiant whiteDwarf
  BrownDwarf          15          0            0        0          0          0
  Hypergiant           0         15            0        0         17          0
  MainSequence         0          0           10        0          0          0
  RedDwarf             0          0            0        9          0          0
  Supergiant           0          0            0        0          0          0
  whiteDwarf           0          0            0        0          0          9
```

Decision Tree - Risultati

	Class: BrownDwarf	Class: Hypergiant
Precision	1.0	0.4688
Recall	1.0	1.0000
F1	1.0	0.6383
Prevalence	0.2	0.2000
Detection Rate	0.2	0.2000
Detection Prevalence	0.2	0.4267
Balanced Accuracy	1.0	0.8583
	Class: MainSequence	Class: RedDwarf
Precision	1.0000	1.00
Recall	1.0000	1.00
F1	1.0000	1.00
Prevalence	0.1333	0.12
Detection Rate	0.1333	0.12
Detection Prevalence	0.1333	0.12
Balanced Accuracy	1.0000	1.00
	Class: Supergiant	Class: WhiteDwarf
Precision	NA	1.00
Recall	0.0000	1.00
F1	NA	1.00
Prevalence	0.2267	0.12
Detection Rate	0.0000	0.12
Detection Prevalence	0.0000	0.12
Balanced Accuracy	0.5000	1.00

Macro:
precision: 0.7447917 - **recall:** 0.8333333 - **f-measure:** 0.7730496 - **AUC:** 0.9519066

Random Forest

**Features:** all

mtry: 11; calcolato da 10-fold cv

```
            Reference
Prediction    BrownDwarf Hypergiant MainSequence RedDwarf Supergiant WhiteDwarf
  BrownDwarf          15          0            0        0          0          0
  Hypergiant           0         15            0        0          0          0
  MainSequence         0          0           10        0          0          0
  RedDwarf             0          0            0        9          0          0
  Supergiant           0          0            0        0         17          0
  whiteDwarf           0          0            0        0          0          9
```

	Class: BrownDwarf	Class: Hypergiant
Precision	1.0	1.0
Recall	1.0	1.0
F1	1.0	1.0
Prevalence	0.2	0.2
Detection Rate	0.2	0.2
Detection Prevalence	0.2	0.2
Balanced Accuracy	1.0	1.0

	Class: MainSequence	Class: RedDwarf
Precision	1.0000	1.00
Recall	1.0000	1.00
F1	1.0000	1.00
Prevalence	0.1333	0.12
Detection Rate	0.1333	0.12
Detection Prevalence	0.1333	0.12
Balanced Accuracy	1.0000	1.00

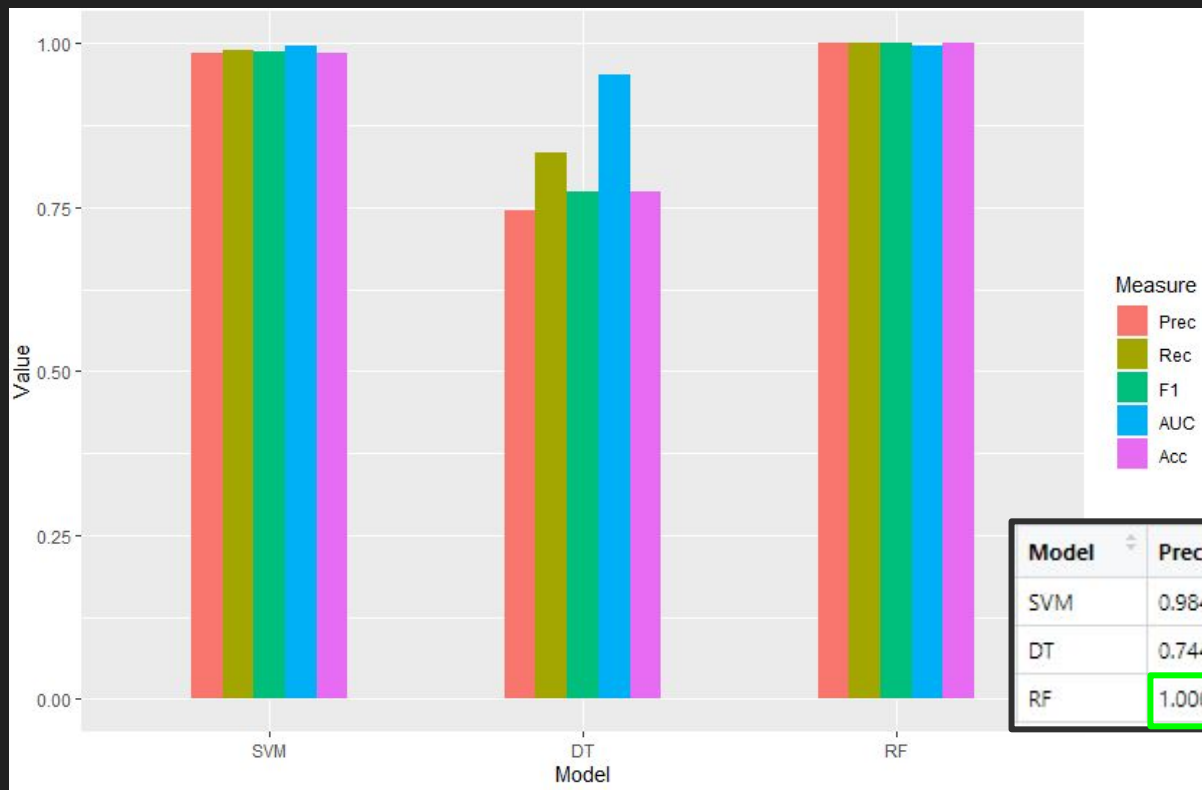|	Class: Supergiant	Class: whiteDwarf
Precision	1.0000	1.00
Recall	1.0000	1.00
F1	1.0000	1.00
Prevalence	0.2267	0.12
Detection Rate	0.2267	0.12
Detection Prevalence	0.2267	0.12
Balanced Accuracy	1.0000	1.00

Macro:

precision: 1 - **recall:** 1

f-measure: 1 - **AUC:** 0.9959891
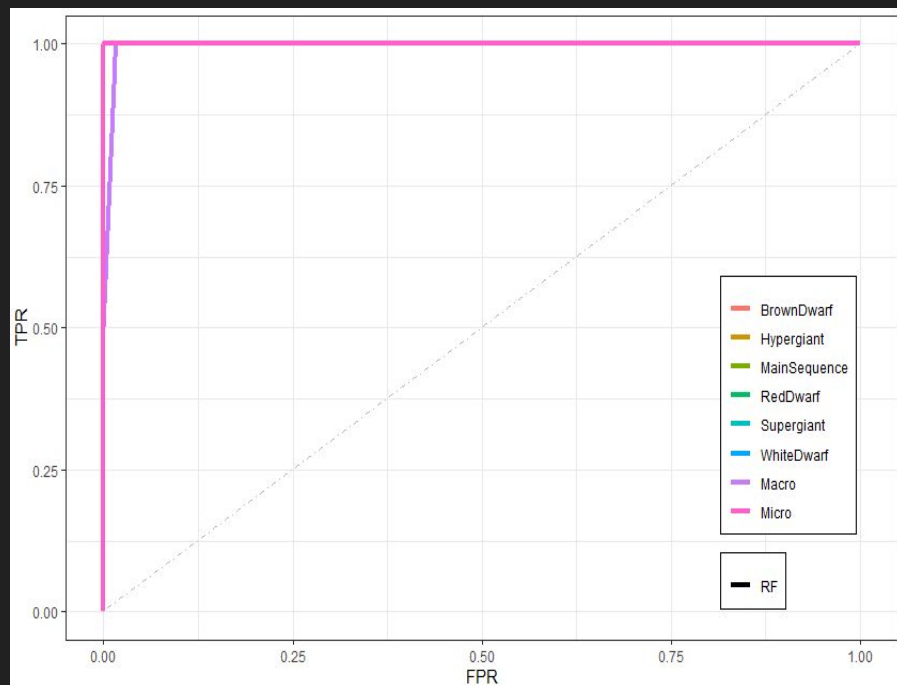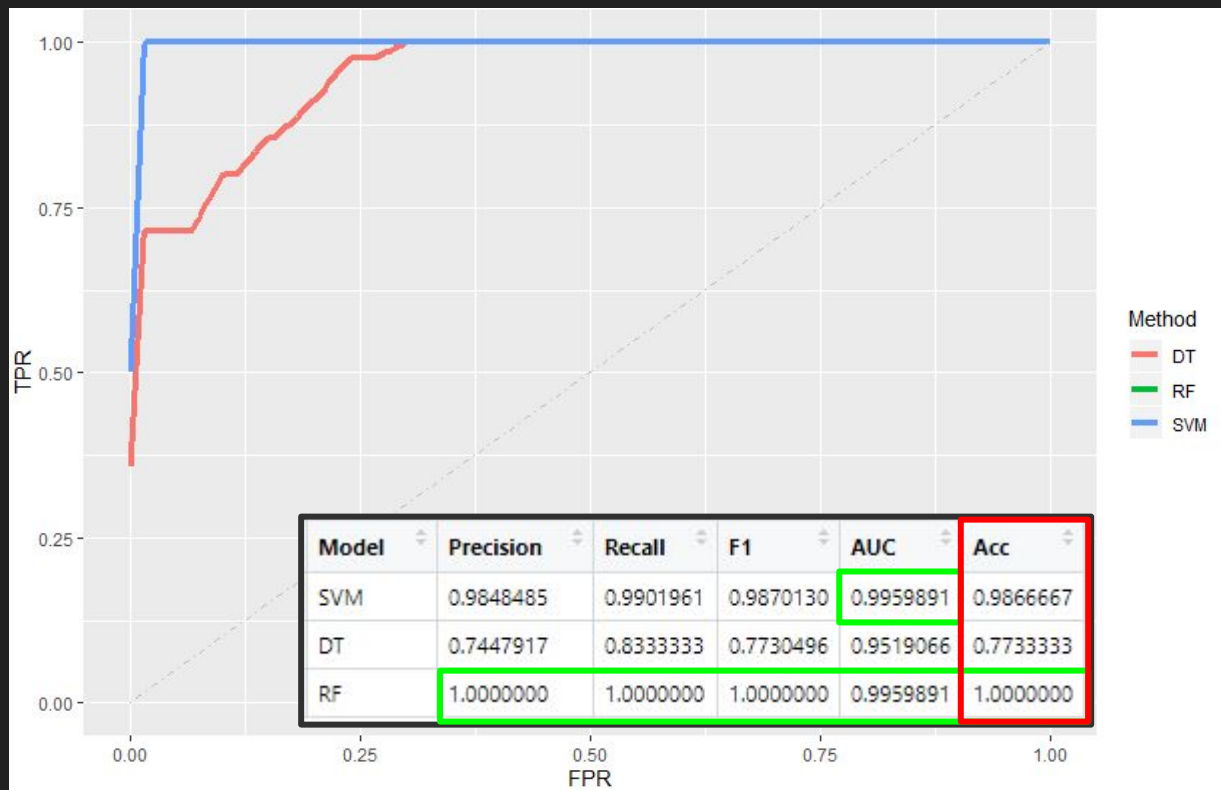
Confronto modelli

Confronto modelli - misure macro

- AUC simile
- DT modello peggiore

Model	Precision	Recall	F1	AUC	Acc
SVM	0.9848485	0.9901961	0.9870130	0.9959891	0.9866667
DT	0.7447917	0.8333333	0.7730496	0.9519066	0.7733333
RF	1.0000000	1.0000000	1.0000000	0.9959891	1.0000000

Confronto modelli - misure macro

- AUC simile
- DT modello peggiore

	Everything	FinalModel
svm	24.25	0.56
dt	15.04	0.02
rf	17.26	0.09

Model	Precision	Recall	F1	AUC	Acc
SVM	0.9848485	0.9901961	0.9870130	0.9959891	0.9866667
DT	0.7447917	0.8333333	0.7730496	0.9519066	0.7733333
RF	1.0000000	1.0000000	1.0000000	0.9959891	1.0000000