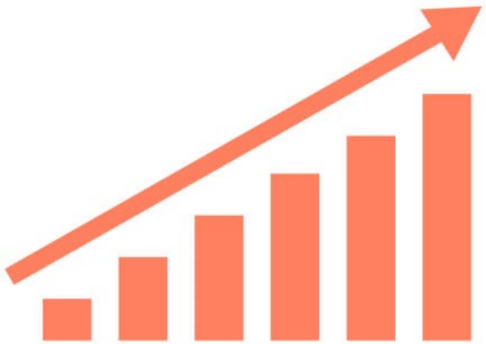
POLITECNICO
MILANO 1863

Numerical Analysis for Machine Learning Project

Group 34
Alberto Sandri
Enrico Simionato

Credit card fraud detection



A machine learning based credit card fraud detection using the GA algorithm for feature selection

Emmanuel Ileberi, Yanxia Sun & Zenghui Wang

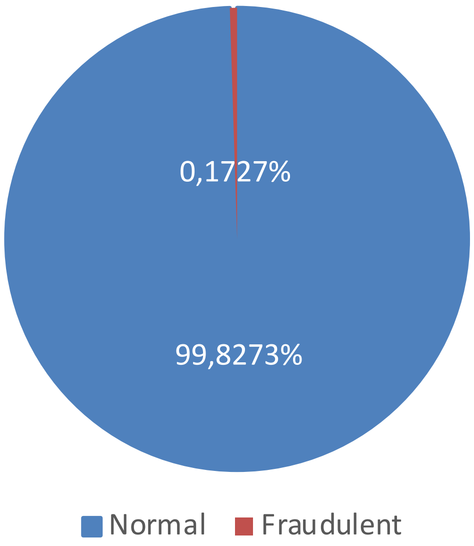
Journal of Big Data **9**, Article number: 24 (2022) | Cite this article

Dataset

Features

Time	V1	V2	…	…	V28	Amount	Class

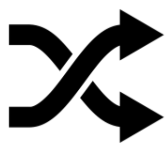
Transactions



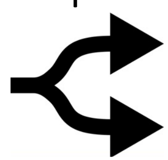
Pre-processing

GA features selection

Shuffle + Split



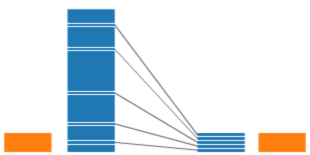
+



Normalization

$$f_{scaled} = \frac{f - min(f)}{max(f) - min(f)}$$

Undersampling



Metrics

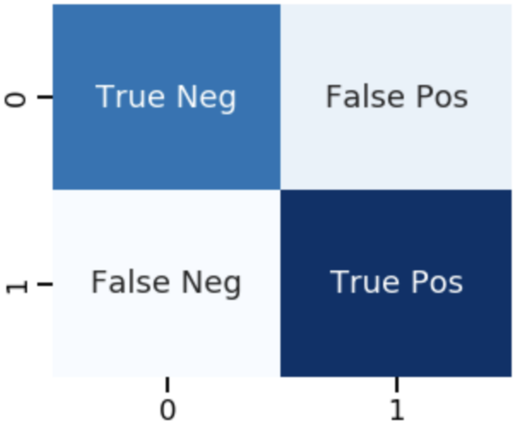
Accuracy

Recall

Precision

F1-score

AUC



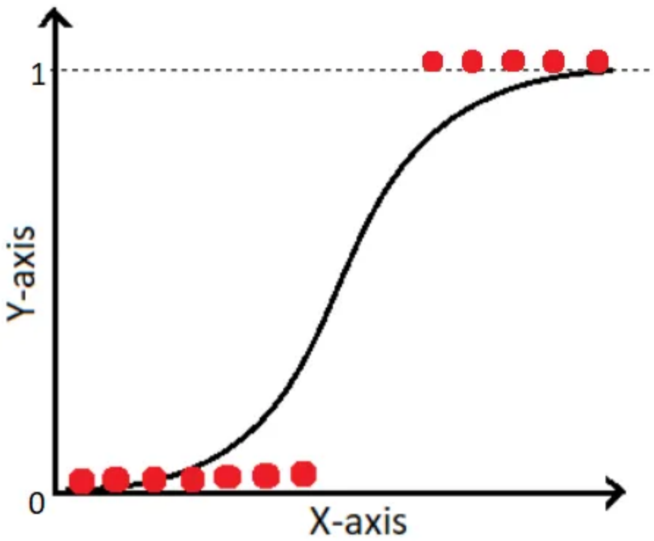
Logistic Regression

Model definition

$$z = w_0x_0 + w_1x_1 + ... + w_nx_n + b$$

$$y_{pred} = \frac{1}{1 + e^{-z}}$$

$$label = \begin{cases} 1, & \text{if } y_{pred} \geq 0.5 \\ 0, & \text{otherwise} \end{cases}$$

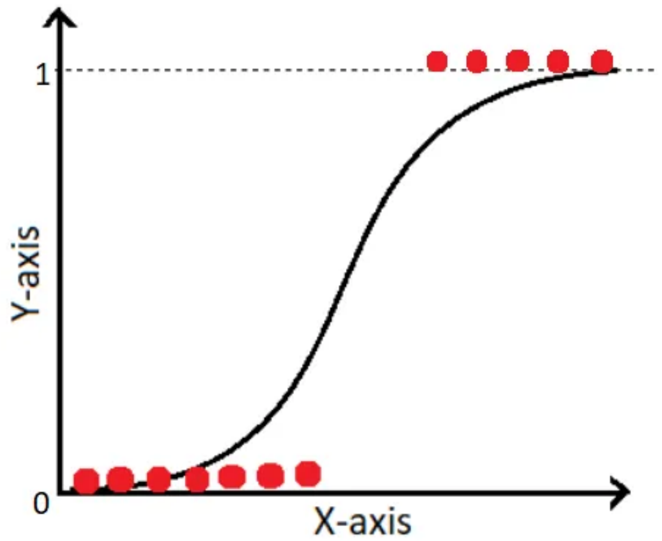


Logistic Regression

$$J(\mathbf{w}, b) = -\frac{1}{n_{samples}} \sum_{i=1}^{n_{samples}} \alpha y_i \log y_{pred,i} + \beta(1 - y_i) \log(1 - y_{pred,i})$$

$$\mathbf{g}(\mathbf{x}^{(k)}) = \frac{1}{|I_k|} \sum_{i_k \in I_k} \nabla J_{i_k}(\mathbf{x}^{(k)})$$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \gamma^{(k)} \mathbf{g}(\mathbf{x}^{(k)})$$



Recall vs precision

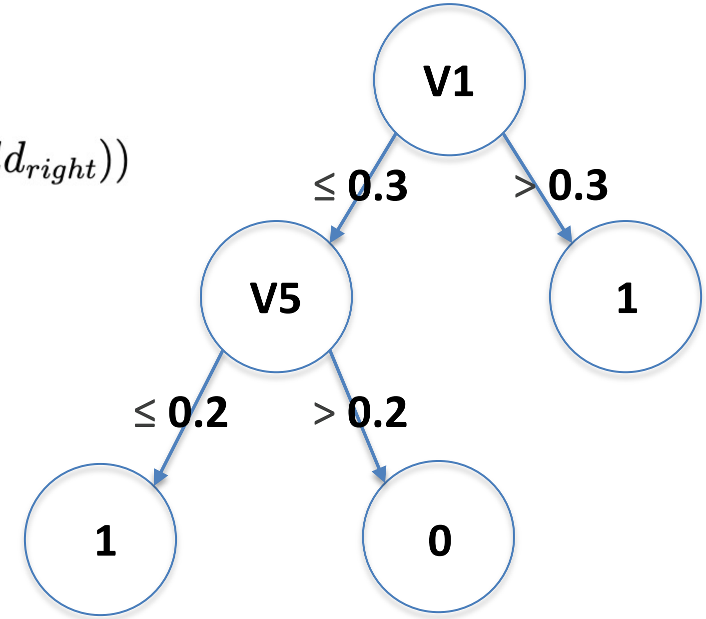
Fast training phase

Decision Tree

Information gain

$$IG = G(parent) - (w_{left} \cdot G(child_{left}) + w_{right} \cdot G(child_{right}))$$

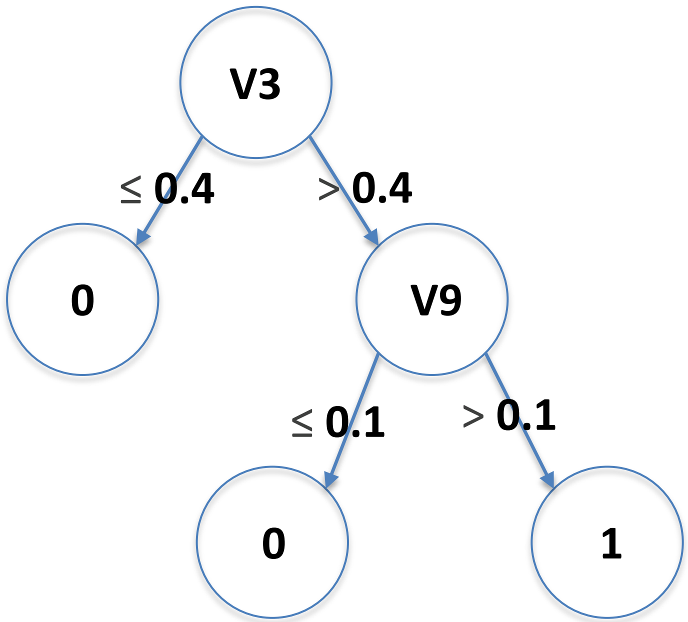
$$G(node) = 1 - \sum_{l \in labels} p(l)^2$$

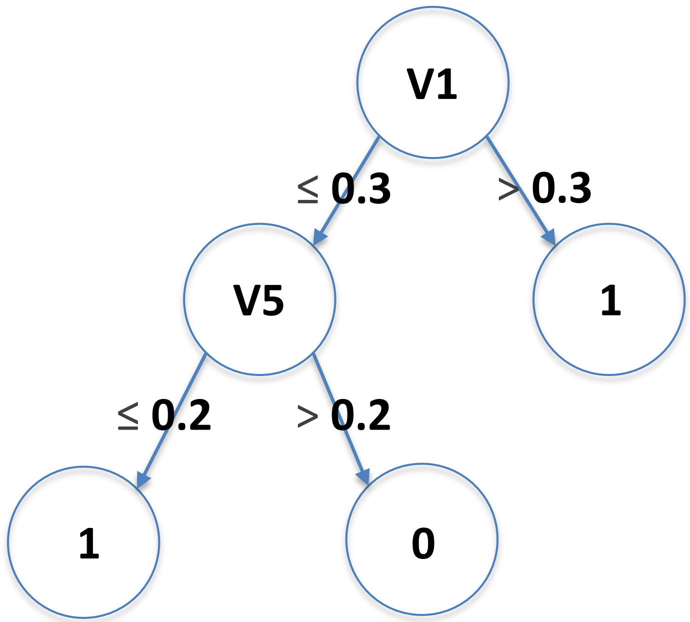


Little time to train

Few hyperparameters to tune

Random Forest





Bootstrapped dataset

Majority vote

Good performances

More memory and training time than DT

Gaussian Naïve Bayes

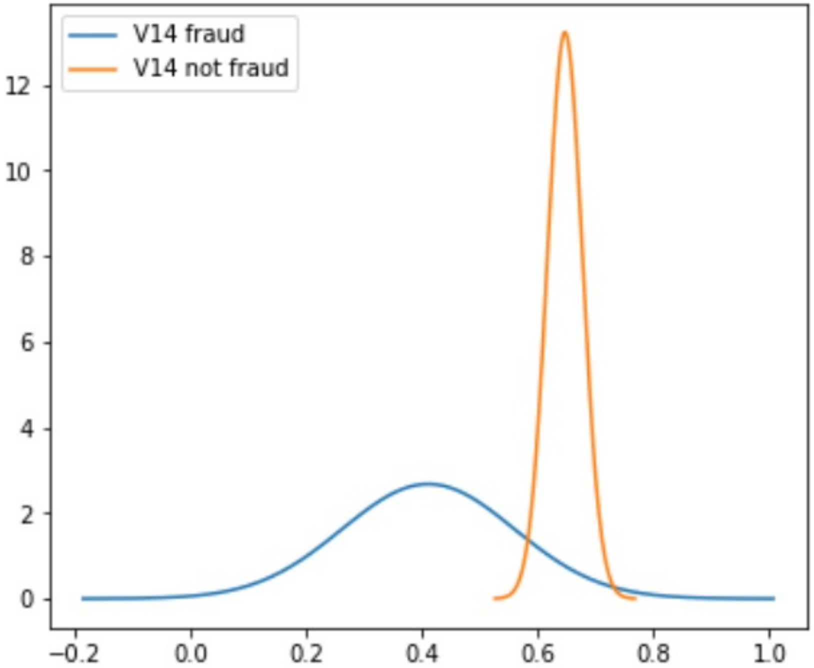
Model definition

Sample mean

$$\overline{X_i} = \frac{1}{n_{samples}} \sum_{j=1}^{n_{samples}} x_{ji}$$

Sample variance

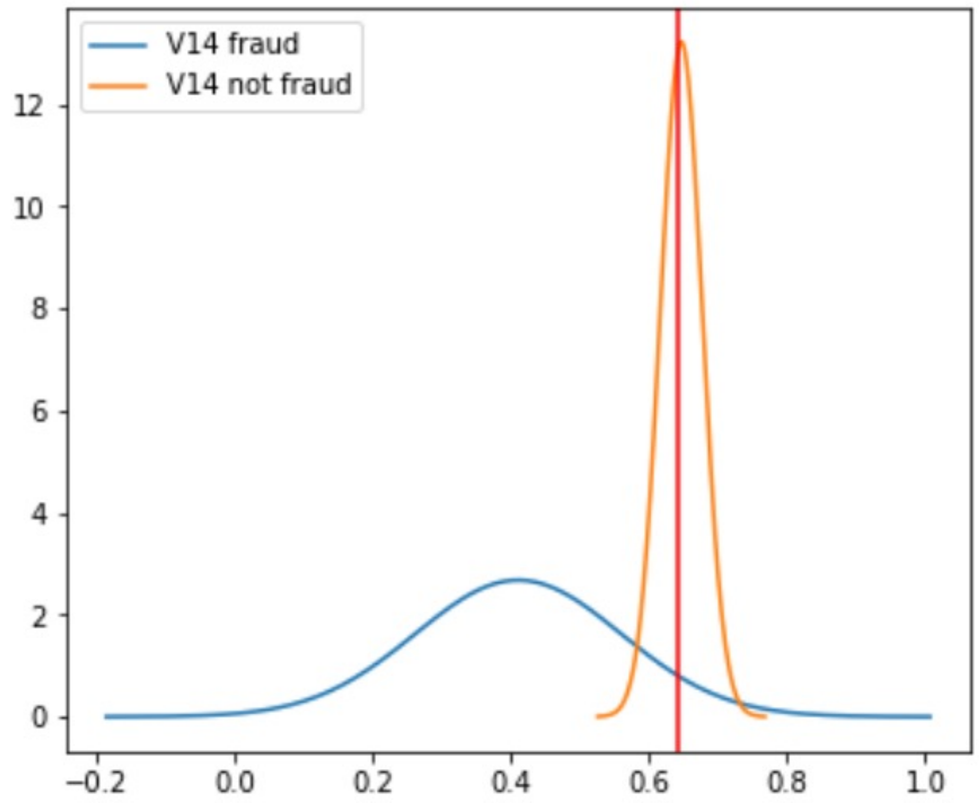
$$var_i = \frac{1}{n_{samples} - 1} \sum_{j=1}^{n_{samples}} (x_{ji} - \overline{X_i})^2$$



$$p(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

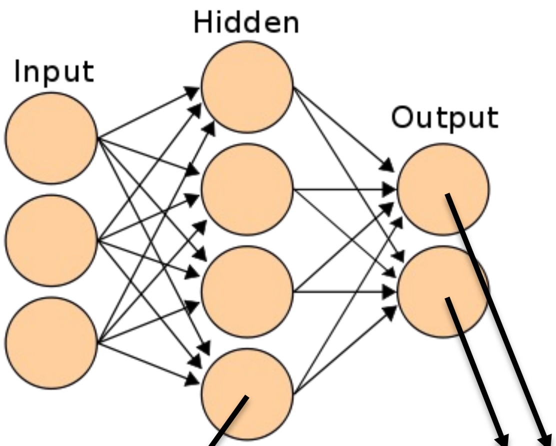
Gaussian Naïve Bayes

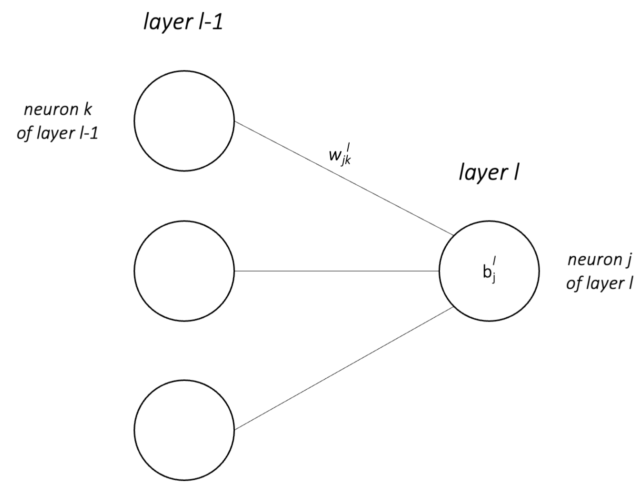
Prediction



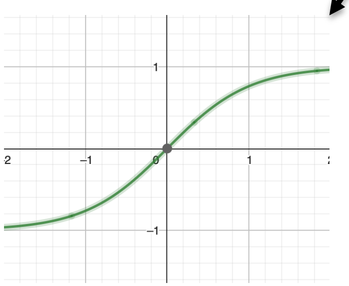
Artificial Neural Network

Neural network schema



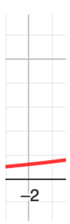


tanh

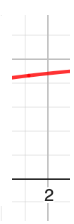


sigmoid

softmax



$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$



$$\mathbf{a}^l = \sigma(\mathbf{W}^l \cdot \mathbf{a}^{l-1} + \mathbf{b}^l)$$

Artificial Neural Network

- **Cost functions**

Cross entropy for 2 classes

$$J(\mathbf{W},\mathbf{b}) = -\frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \alpha y(\mathbf{x}_i) \log a_i^L + \beta(1-y(\mathbf{x}_i)) \log\left(1-a_i^L\right)$$

Cross entropy for n_{output} classes

$$J(\mathbf{W},\mathbf{b}) = -\frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \sum_{j=0}^{n_{outputs}-1} \alpha_j y_j(\mathbf{x}_i) \log a_{ji}^L$$

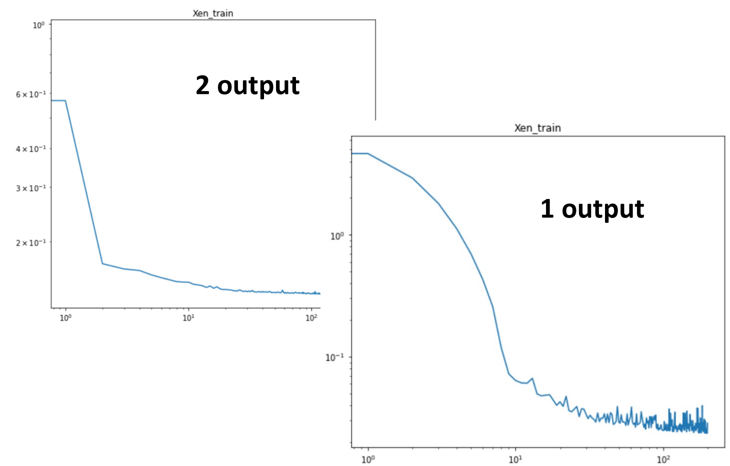
- **Optimization methods**

$$\mathbf{g}(\mathbf{x}^{(k)}) = \frac{1}{|I_k|} \sum_{i_k \in I_k} \nabla J_{i_k}(\mathbf{x}^{(k)})$$

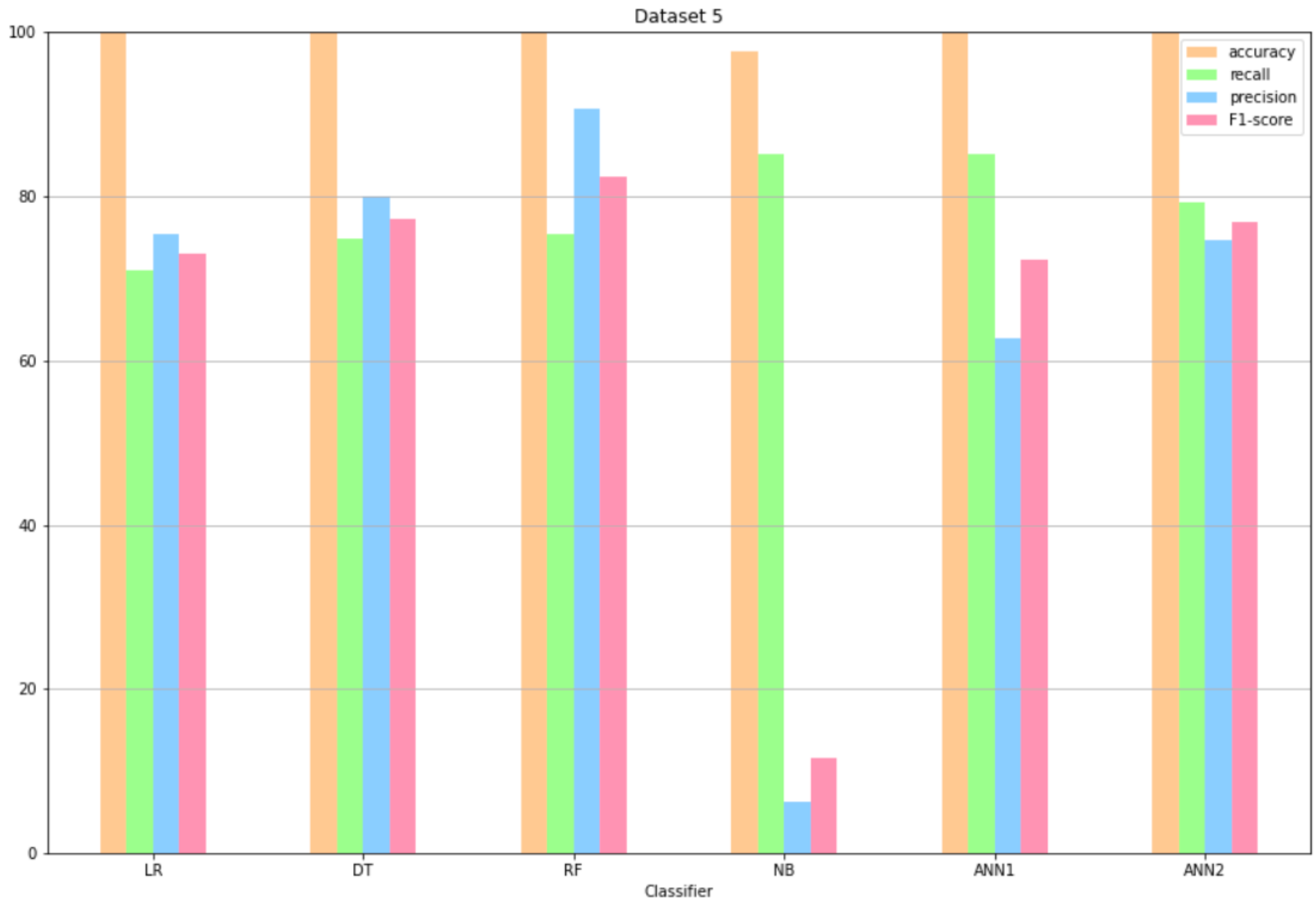
RMSprop

$$\mathbf{r}^{(k+1)} = \rho \mathbf{r}^{(k)} + (1-\rho)\mathbf{g}(\mathbf{x}^{(k)}) \odot \mathbf{g}(\mathbf{x}^{(k)})$$

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - \frac{\lambda}{\delta + \sqrt{\mathbf{r}^{(k+1)}}} \odot \mathbf{g}(\mathbf{x}^{(k)})$$



Results

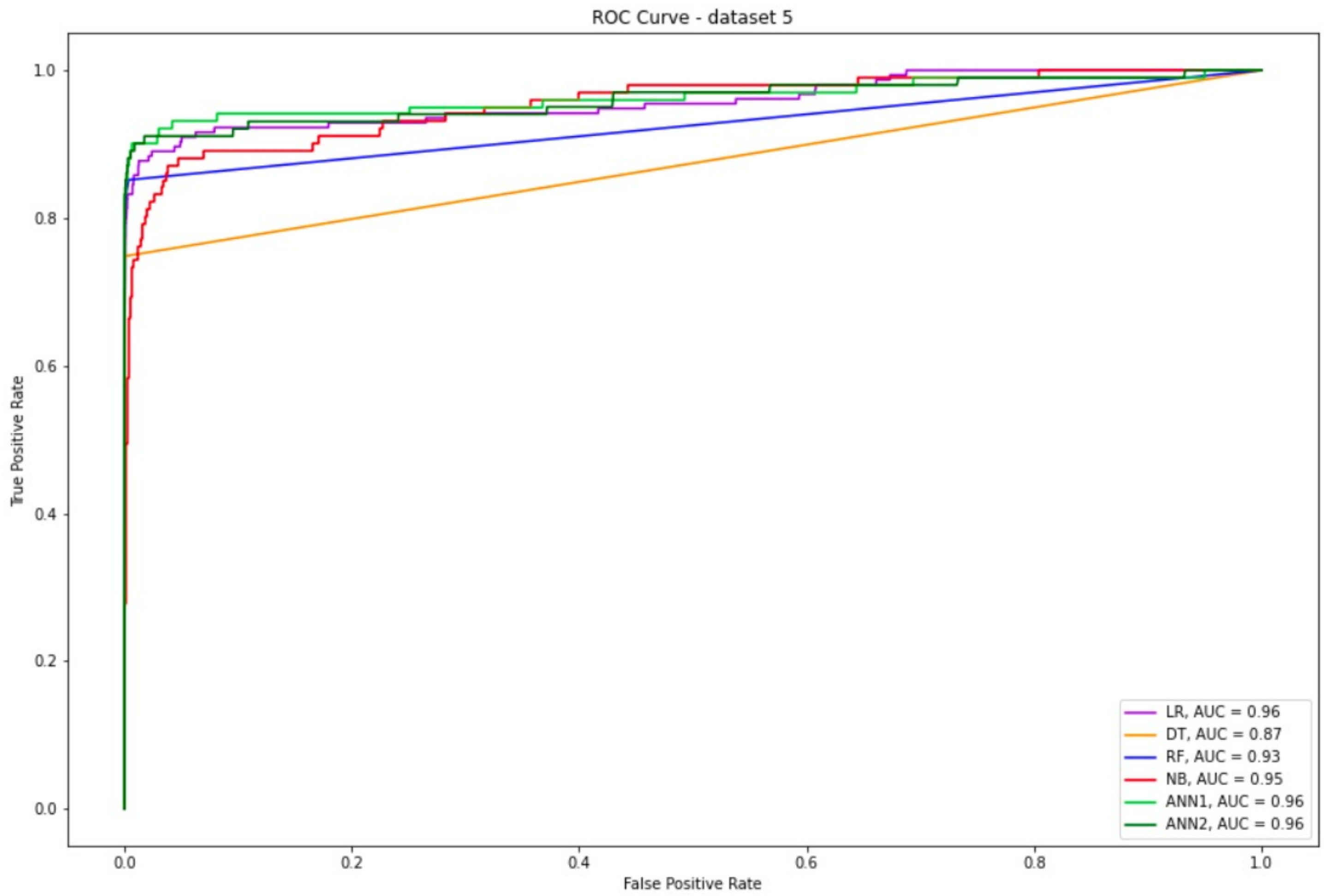


Dataset 5
100
80
60
40
20
0
accuracy
recall
precision
F1-score
LR
DT
RF
NB
ANN1
ANN2
Classifier

Results

Model	Accuracy	Recall	Precision	F1-Score
RF	99.94	75.48	90.70	82.39
	99.98	72.56	95.34	82.41
DT	99.92	74.84	80.00	77.33
	99.89	72.56	65.07	68.61
ANN1	99.88	85.15	62.77	72.27
ANN2	99.92	79.21	74.77	76.92
	99.08	77.87	12.27	21.20
NB	97.69	85.15	6.20	11.57
	99.44	57.52	15.85	24.85
LR	99.91	70.97	75.34	73.09
	99.77	46.90	34.64	39.84

Results



ROC Curve - dataset 5
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
LR, AUC = 0.96
DT, AUC = 0.87
RF, AUC = 0.93
NB, AUC = 0.95
ANN1, AUC = 0.96
ANN2, AUC = 0.96

Conclusion

Overall good performances

Using different hyperparameters for the different datasets could improve performaces

GA feature selection

Conclusion

Accuracy is very important but the recall is the keypoint

VS

Tradeoff recall – precision

Hyperparameters tuning is challenging