



UNIVERSITÀ DEGLI STUDI DI SALERNO

Dipartimento di Scienze Economiche e Statistiche
Corso di Laurea Triennale in Statistica Per i Big Data

MACHINE LEARNING FOR FRAUD DETECTION

Prova finale in
Statistical Learning

RELATORE:

CH.MO PROF. PIETRO CORETTO

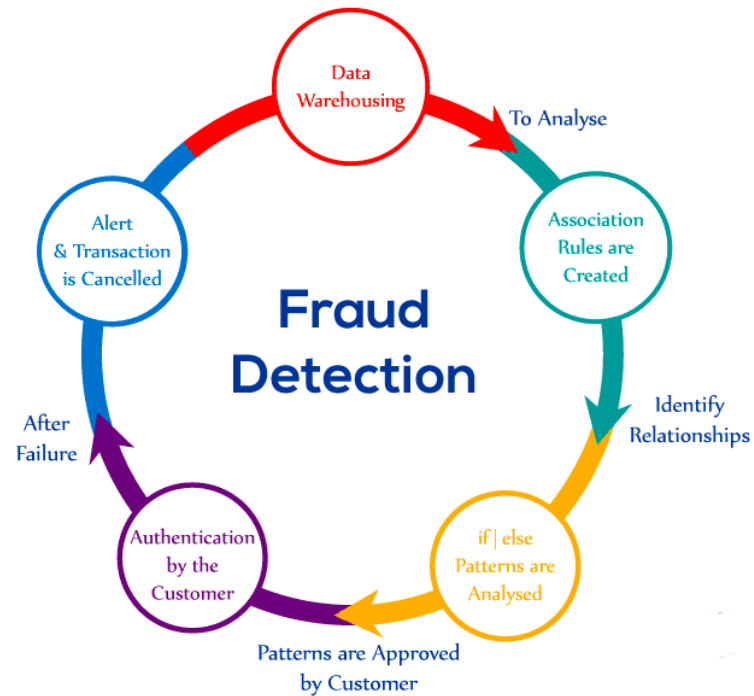
CANDIDATO:

DAVIDE MASCOLO

MATR. 0212800247

ANNO ACCADEMICO 2020/2021

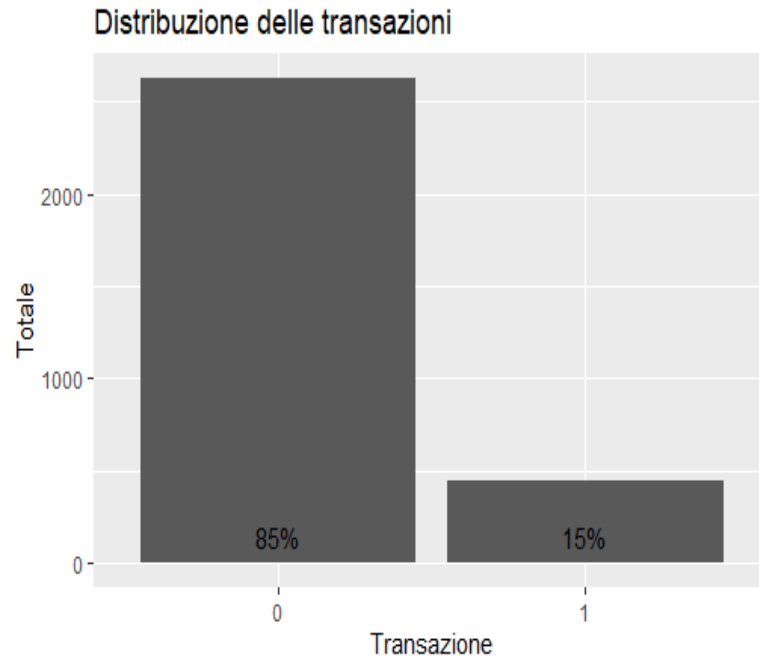
FRAUD DETECTION



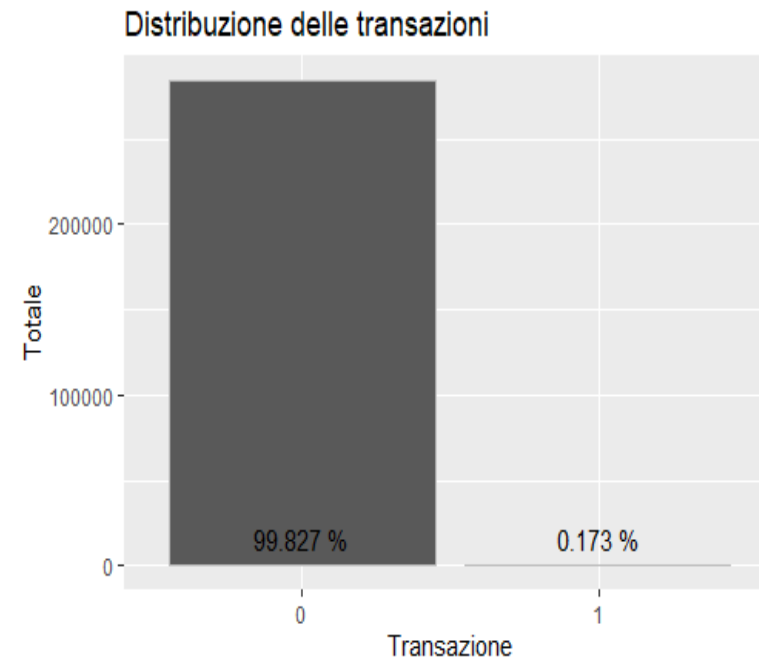
«Insieme dei processi ed analisi che consentono di identificare e prevenire attività finanziarie non autorizzate per impedire l'ottenimento di denaro tramite falsi pretesti»

Alexander S. Gillis, 2019

UNBALANCED CLASSIFICATION



Caso Studio 1: Small Card Data



Caso Studio 2: European Card Data

ACCURACY PARADOX

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Confusion Matrix

Sensitivity

Specificity

Precision

F_β - Score

Matthews Correlation Coefficient

Cohen's Kappa

Receiver Operating Characteristic (ROC)

METODI

REGRESSIONE LOGISTICA

ANALISI DISCRIMINANTE LINEARE

ANALISI DISCRIMINANTE QUADRATICA

ALBERO DI DECISIONE

APPLICAZIONE CON DATI REALI

Caso 1: Small Card Data

TOTALE TRANSAZIONI	3075
TRANSAZIONI NON FRAUDOLENTE	2627
TRANSAZIONI FRAUDOLENTE	448
% NON FRAUDOLENTE	85,431
% FRAUDOLENTE	14,569

RISULTATI

REGRESSIONE LOGITSTICA

CONFUSION MATRIX (KFCV)		
Actual Values		
Predicted Values	0	1
0	85,17 %	0,6 %
1	0,24%	13,99 %

CONFUSION MATRIX (EXTERNAL)		
Actual Values		
Predicted Values	0	1
0	261	1
1	1	43

EVALUATION METRICS		
	KFCV	EXTERNAL
Sensitivity	0,959	0,977
Specificity	0,958	0,996
Precision	0,984	0,977
F_1 Score	0,970	0,977
F_2 Score	0,963	0,977
Mcc	0,966	0,973
KCohen	0,966	0,974
Auc	0,978	0,987



APPLICAZIONE CON DATI REALI

Caso 2: European Card Data

TOTALE TRANSAZIONI	284807
TRANSAZIONI NON FRAUDOLENTE	284315
TRANSAZIONI FRAUDOLENTE	492
% NON FRAUDOLENTE	99,827
% FRAUDOLENTE	0,173

RISULTATI

ANALISI DISCRIMINANTE LINEARE

CONFUSION MATRIX (KFCV)		
Actual Values		
Predicted Values	0	1
0	99,81 %	0,04 %
1	0,02 %	0,13 %

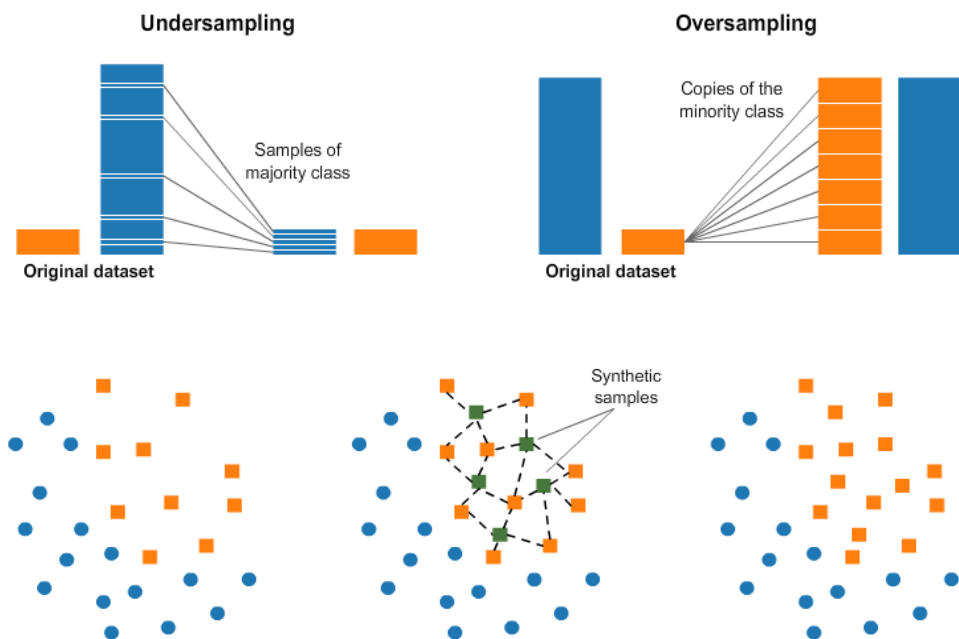
CONFUSION MATRIX (EXTERNAL)		
Actual Values		
Predicted Values	0	1
0	28421	11
1	10	38

EVALUATION METRICS		
	KFCV	EXTERNAL
Sensitivity	0,765	0,776
Specificity	0,999	0,999
Precision	0,876	0,792
F_1 Score	0,815	0,784
F_2 Score	0,784	0,779
Mcc	0,820	0,820
KCohen	0,814	0,783
Auc	0,882	0,888

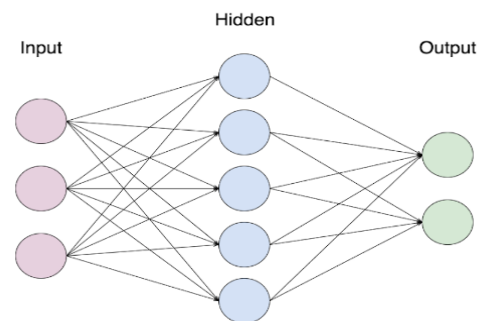
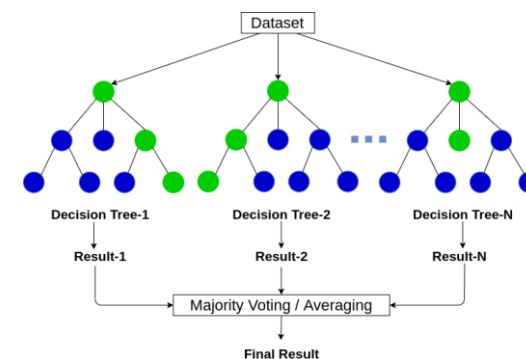
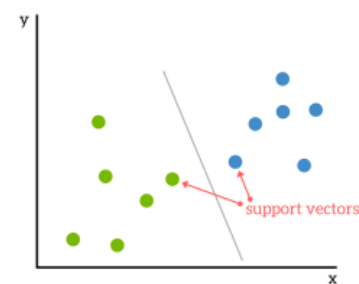


CONCLUSIONI

RE - SAMPLING



ALTRI METODI



GRAZIE PER L'ATTENZIONE



PROVA FINALE IN STATISTICAL LEARNING - MACHINE LEARNING FOR FRAUD DETECTION

