**Academic Year:** 2023/2024

**Course:** Artificial Intelligent Systems - Intelligent Model Project

**Prof. Alfredo Milani**

**Ongoing assignmen**t: Machine Learning Basics

**Student:** Lorenzo Mariotti

**ID:** 369094

Sommario

[1. Dataset 3](#_Toc154788362)

[2. Classification 4](#_Toc154788363)

[2.1. Decision Tree 5](#_Toc154788364)

[2.2. Nearest Neighbor 6](#_Toc154788365)

# Dataset

**Link:**

<https://www.kaggle.com/datasets/fedesoriano/company-bankruptcy-prediction/data>

**Descrizione:**

Il dataset riporta i dati sulla bancarotta di aziende tratti dal *Taiwan Economic Journal* per gli anni 1999-2009. Il fallimento dell'azienda è stato definito in base al regolamento aziendale della Borsa di Taiwan.

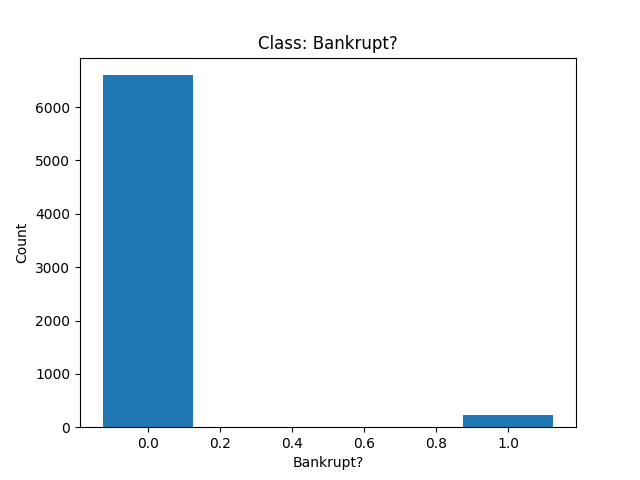
**Attributi:**

Il dataset non presenta alcun attributo categorico quindi non è necessario alcuna fattorizzazione.

**Classi:**

Le classi d’interesse sono rappresentate dalla colonna *“Bankrupt?”* che riporta con valori booleani se una data condizione ha portato o meno alla bancarotta dell’azienda.

**Distribuzione delle classi:**



* Numero di classi “0” = **6599**
* Numero di classi “1” = **220**

Uno sbilanciamento tale di classi rende metriche come l’accuratezza forvianti.

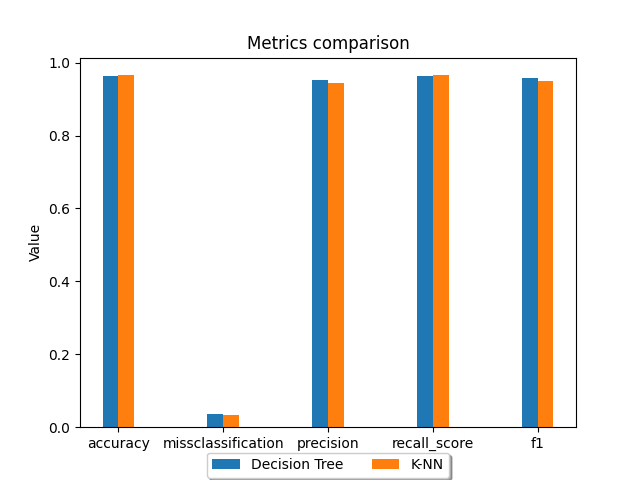
# Classification

**Dimensione del training set:** 75%

**Dimensione del testing set:** 25%

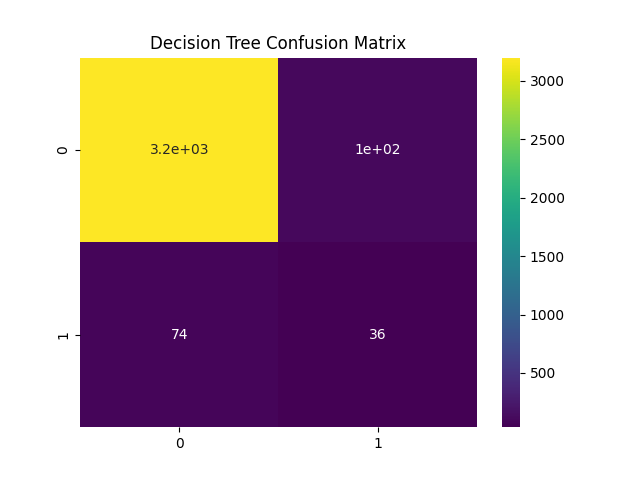
**Classificatori analizzati:**

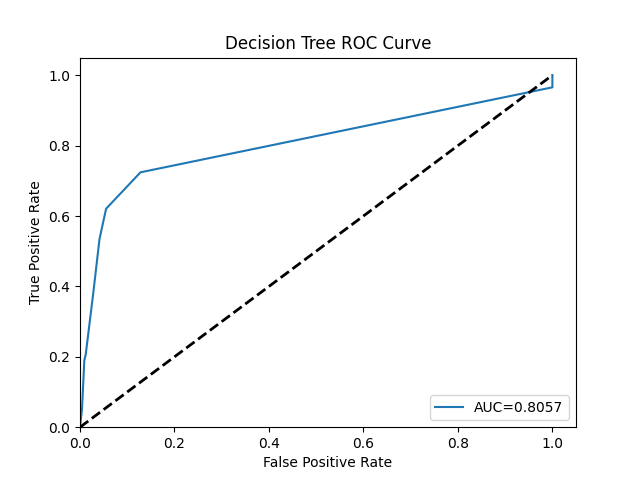
* Decision Tree
* Nearest-neighbor



## Decision Tree

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Accuracy | Missclassification | Precision | Recall score | F1 |
| 0. 963636 | 0. 036364 | 0. 953334 | 0. 963636 | 0. 956885 |





## Nearest Neighbor

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Accuracy | Missclassification | Precision | Recall score | F1 |
| 0. 965396 | 0. 034604 | 0. 944971 | 0. 965396 | 0. 95008 |

