# Riphah International University
(Main Campus)

## DATA SCIENCE

## Mr. Imran Khan

# SEMESTER PROJECT

# Muhammad Hammad Butt
# SAP Id: 59944

# Stock Market Prediction

## Introduction

The stock market is inherently volatile and influenced by numerous factors, making accurate trend prediction a challenging task. This project aims to utilize RapidMiner, a powerful data science platform, to develop a predictive model that can forecast stock market trends based on historical trading data.

## Dataset Description

The dataset used in this project is the stock exchange market Dataset from the UCI Machine Learning Repository, available on Kaggle. It contains various attributes describing the fluctuations, such as market conditions at different points.
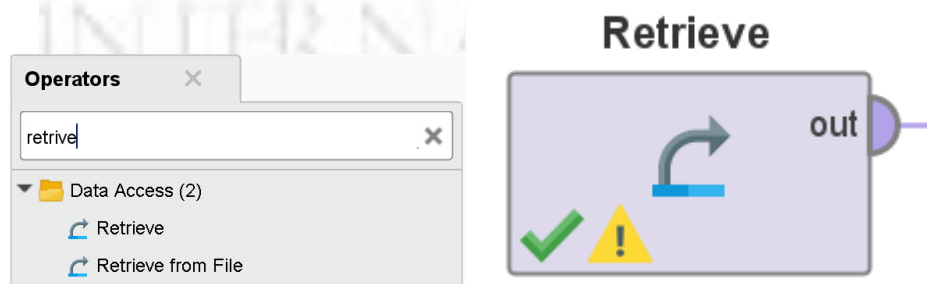
- Number of instances: 3900
- Number of attributes: 8 (e.g., Date, Open, High, Low, Close, Volume, Ticker)
- Data type: Categorical

## Tools and Software Used

- RapidMiner Studio: For data processing, model building, and evaluation
- Microsoft Word: For documenting the project
- Operating System: (Windows 11pro)
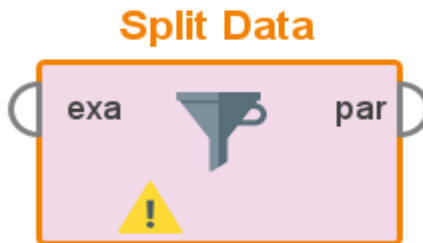- Implementation Plan

## Data Import and Preparation

- Imported the stock market dataset using the **Retrieve** operator in RapidMiner.



- Set the **role** of the target attribute (class) as label using the **Set Role** operator.

- Split the data into training and testing sets using the **Split Data** operator with a 70:30 ratio.
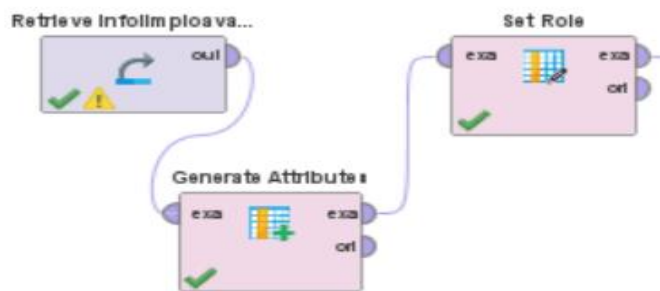
**Split Data**

exa ◉ par

⚠️

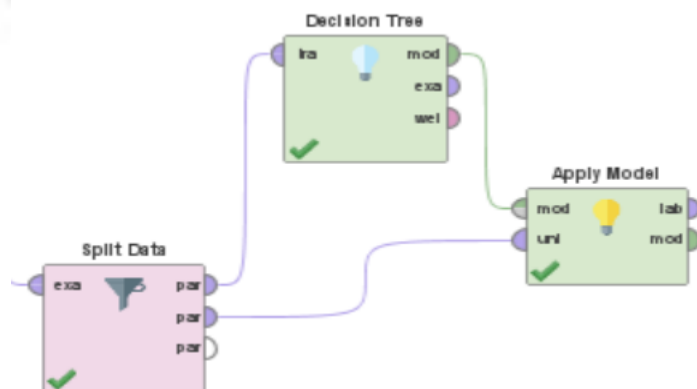- Used stratified sampling to maintain the distribution of classes in both datasets.

## **Process DesignClass Definition:**

The process workflow in RapidMiner includes:
- **Retrieve:** Loads the stock market dataset.
- **Set Role:** Assigns the target variable as the label

Retrieve Infolimploava...    Set Role
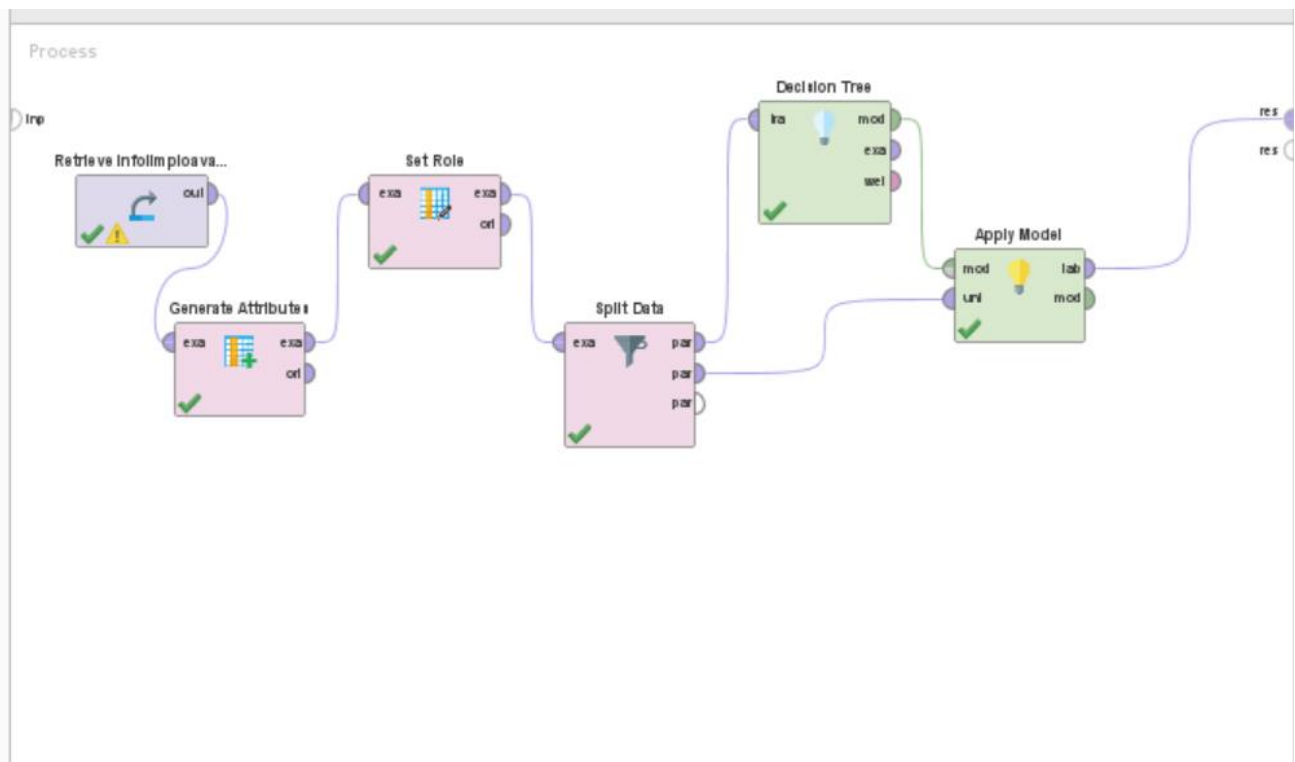
Generate Attributes

- .
- **Split Data:** Splits data into training and testing sets.
- **Decision Tree:** Builds a classification model using training data.
- **Apply Model:** Applies the trained model on the test data.
- **Performance:** Evaluates model accuracy and confusion matrix.

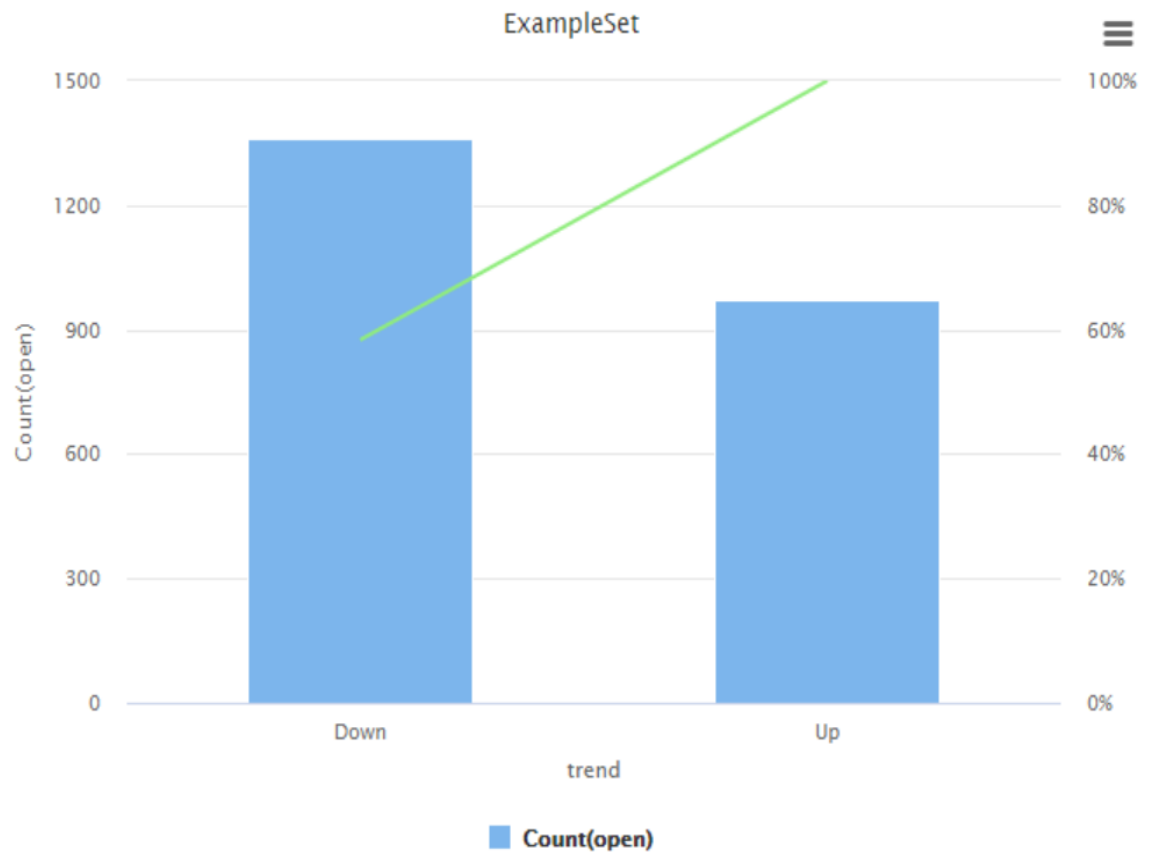Decision Tree

Apply Model

Split Data

## Model Building and Evaluation:

- Used the Decision Tree algorithm for classification due to its interpretability and effectiveness with categorical data.
- Evaluated the model with metrics such as accuracy and confusion matrix.
- Achieved an accuracy of approximately 99.9%, indicating excellent predictive performance.

# <u>Visualization:</u>



-

- 

  - (Graphical View)
    - 

| | % PerformanceVector (Performance) | ✕ |
|---|---|---|

# PerformanceVector

```
PerformanceVector:
accuracy: 99.90%
ConfusionMatrix:
True:     p         e
p:        3908      0
e:        8         4208
```

## Results and Analysis:

- Steps: The confusion matrix showed very few misclassifications between edible and poisonous mushrooms.
- High accuracy suggests the model can reliably classify mushrooms based on their features.
- This model can be used as a decision support tool for mushroom identification.
- 

## Conclusion:

The Mushroom Classification project successfully demonstrates how machine learning can be applied to real-world data for classification tasks. Using RapidMiner, the decision tree model efficiently classified mushrooms with high accuracy, showcasing the value of data science in safety-critical applications

## References:

- UCI Machine Learning Repository: Mushroom Dataset
- RapidMiner Documentation: https://docs.rapidminer.com
- Kaggle: **https://www.kaggle.com/datasets/uciml/mushroom-classification**