

REPORT

WARDA SHAKEEL
SEC "B"
ROLL NO :19122090
SEMESTER 6TH

"Predicting Educational Domain for Fresh Intermediate Graduates"

OBJECTIVE:

The primary objective of this project is to leverage a dataset collected from university students to predict the preferred educational domain or field of interest for fresh students who have recently passed their intermediate education. By analyzing the historical data of current students' interests and satisfaction with their chosen educational domains during their intermediate years, we aim to create a predictive model that assists in guiding new students towards suitable educational paths.

INTRODUCTION AND BACKGROUND OF THE PROBLEM:

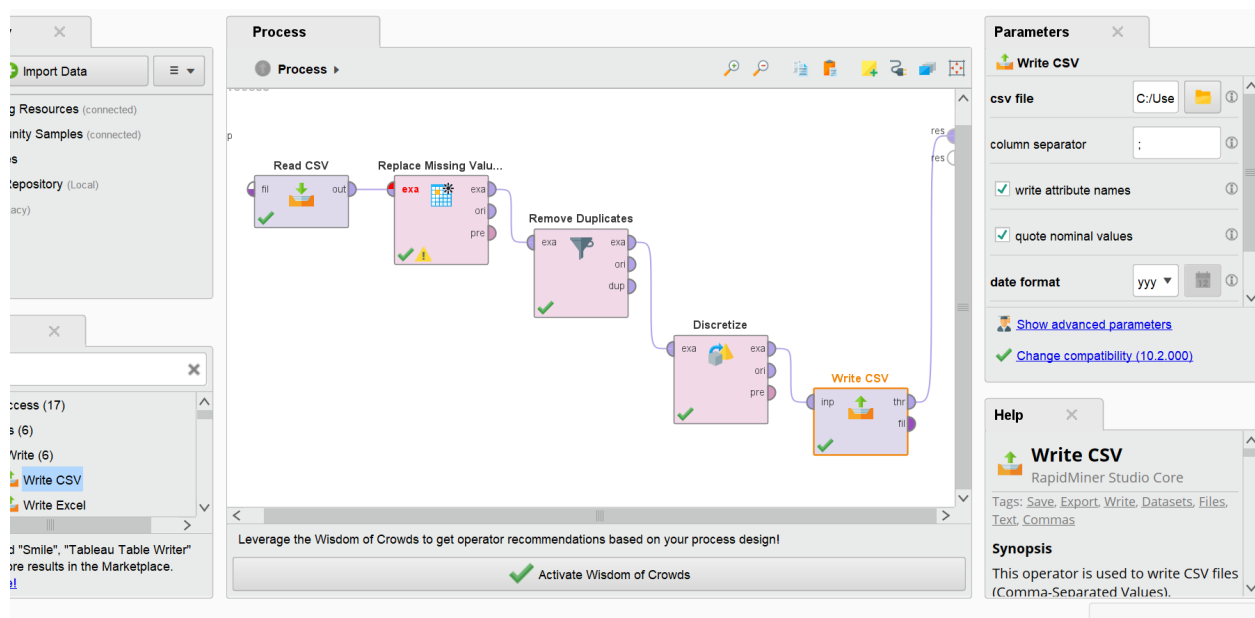
There is an issue where new students often struggle to understand which field they should pursue, leading to them making incorrect decisions that cause problems later on. To solve this problem, a model has been predict to provide them with some guidance in choosing their domain according to their Interest.

COLLECT DATASET FROM WHERE:

This dataset is collected from university students or college students to get their interest when they are in intermediate and get to know that they are happy with their decisions regarding the choice of educational domain.

DATA PREPROCESSING:

Preprocessing often leads to better model performance by making data more suitable for the algorithms, enhancing their ability to learn patterns and make accurate predictions or classifications. I applied preprocessing because there was a noise in the data, resulting in poor outcomes. After preprocessing, significant changes occurred in the dataset. Extra rows were removed, and the data became smoother.



MODELLING AND EVALUATION:

The data seems to have categorical elements like favorite subjects, major subjects in different levels of education, and areas of interest, classification algorithms like decision trees, random forests could be applied. They could predict or classify a student's likely field of study based on their educational background and interests. Students have various subject preferences, such as Biology, Computer Science, Mathematics, Physics, Finance/Accounting/Statistics, etc., indicating a diverse range of academic interests.

Process

Process

```

graph LR
    Inp((inp)) --> ReadCSV[Read CSV]
    ReadCSV --> SplitData[Split Data]
    SplitData --> RandomForest[Random Forest]
    SplitData --> ApplyModel[Apply Model]
    RandomForest --> ApplyModel
    ApplyModel --> Performance[Performance]
  
```

Parameters

Random Forest

- number of trees: 100
- criterion: gain_ratio
- maximal depth: 10
- ☐ apply pruning
- ☐ apply prepruning
- [Show advanced parameters](#)
- [Change compatibility \(10.2.000\)](#)

Help

Random Forest

Concurrency

Tags: Supervised, Classification, Regression, Model, Ensembles, Decision Trees, Extremely Randomized Trees, Extra-Trees, Breiman, Bagging

Synopsis

Leverage the Wisdom of Crowds to get operator recommendations based on your process design!

☒ Activate Wisdom of Crowds

Process

Process

```

graph LR
    Inp((inp)) --> ReadCSV[Read CSV]
    ReadCSV --> SplitData[Split Data]
    SplitData --> DecisionTree[Decision Tree]
    SplitData --> ApplyModel[Apply Model]
    DecisionTree --> ApplyModel
    ApplyModel --> Performance[Performance]
  
```

Parameters

Performance (Performance (Classification))

- main criterion: first
- ☒ accuracy
- ☐ classification error
- ☐ kappa
- ☐ weighted mean recall
- [Show advanced parameters](#)

Help

Performance (Classification)

RapidMiner Studio Core

Tags: Accuracy, Errors, Precision, Recall, K-Squared, Relative, Validations, Evaluations, Metrics, Confusion Matrix, Predictive

Synopsis

Leverage the Wisdom of Crowds to get operator recommendations based on your process design!

☒ Activate Wisdom of Crowds

Time Stamp:Records the date and time of data collection.

Major Subjects in Matric & Intermediate: Specifies the major subjects studied during matriculation (high school) and intermediate education.

Matric & Intermediate Percentage: Indicates the percentage or grade obtained in matriculation and intermediate education.

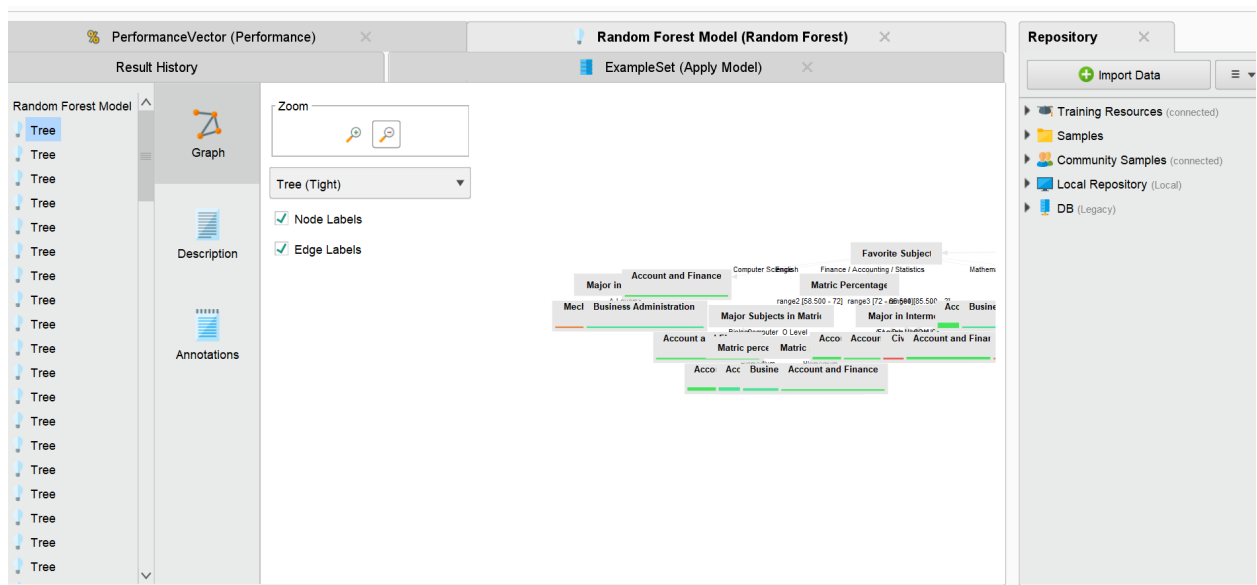
Favorite Subject: Refers to the subject that the student prefers or enjoys the most.

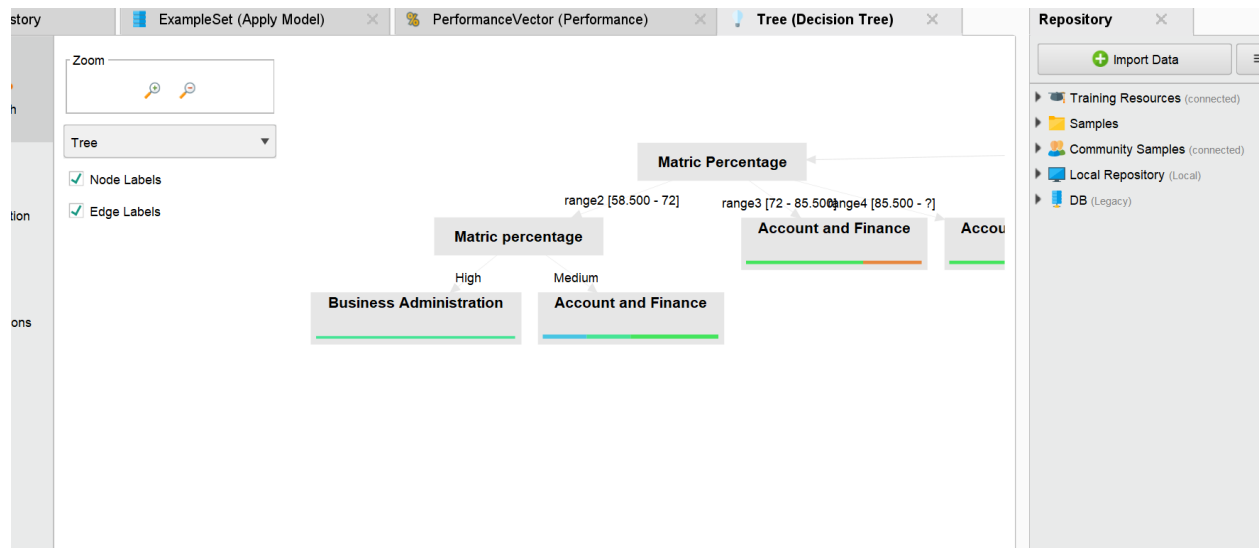
Area of Interest: Specifies the broader area or field that the student finds interesting or has an inclination towards.

Right Domain(Label Column): showcasing the possibility of changing preferences choices in their educational paths.

RESULTS AND CONCLUSION:

I analyzed that the choice of subjects among students heavily influences their preferred domains. For instance, students who favorite subjects like accounting, statistics, or English often option for fields like accounting and finance. Those inclined towards physics or biology tend to choose science-related domains, while students with a background in computer science often apply for IT-related fields. According to the results of the model, more students are choose for three fields in particular: bio science, computer science, and business administration in comparative of other fields.





GITHUB LINK:

<https://github.com/warda1212g/Project-DM.git>