

Bachelor Project

Change detection in event logs to analyze the adapted dataset

Presenter : Siheon Lee

First Supervisor : D. Karastoyanova

Second Supervisor : A. Yadegari Ghahderijani

Outline

- ▶ Introduction
- ▶ Methods
- ▶ Dataset Analysis
- ▶ Implementation
- ▶ Results
- ▶ Conclusion & Future work
- ▶ References

Introduction

Introduction & Motivation

- **Background**
 - **Business Process Management (BPM) & Process Mining (PM)**
 - **Change Detection**
- **Motivation**
 - **Scarcity of Research related to Key Performance Indicators (KPIs)**
- **Expectation**

Research Questions

- 1. What is the difference between raw and adapted datasets?**
- 2. How can a machine learning algorithm, such as a decision tree, be applied to detect changes in the adapted dataset?**
- 3. What are the most influential data attributes when considering intervention as the class attribute?**

Methods

Process

- **Background Research**
- **Data Analysis** (Raw & Adapted dataset)
- **Implementation** (Applying Decision Tree to the adapted dataset)
- **Evaluation** (Compare the results with other ML algorithms)

Tools & Material

- Jupyter Notebook
- Scikit-Learn
- Apromore



- Raw event logs from BPIC 2017
- Adapted event logs from BPIC 2017

Dataset Analysis

Dataset Analysis

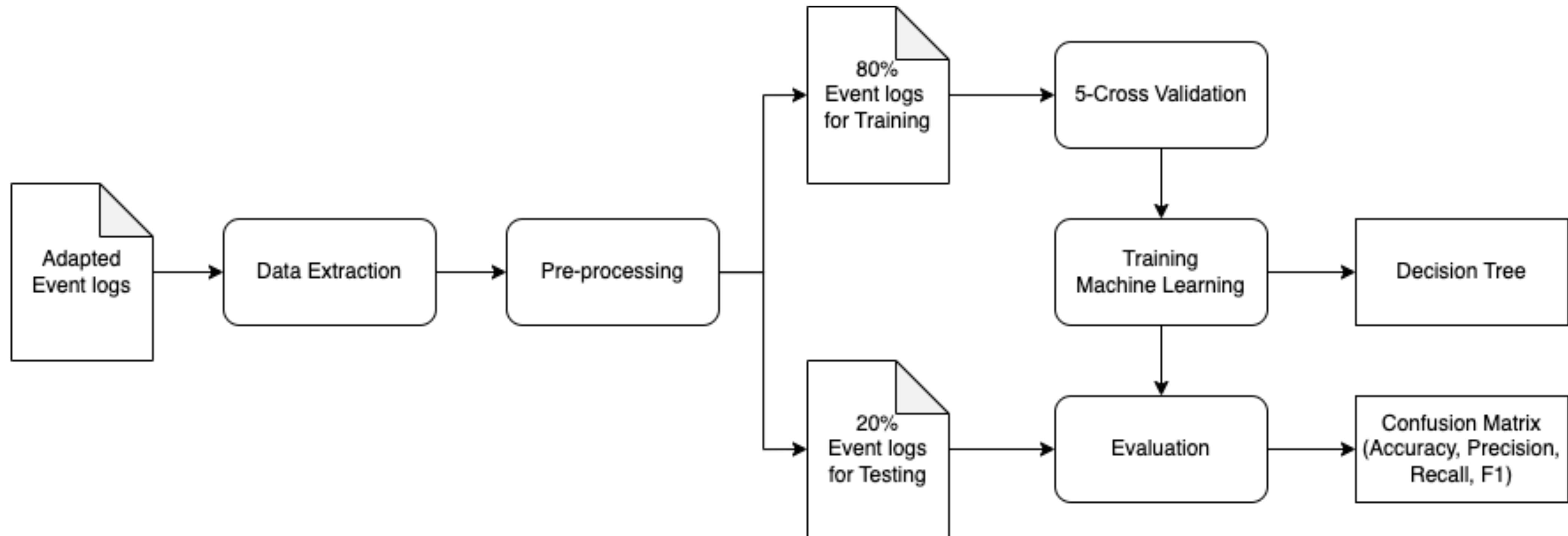
- **Raw dataset**
 - **Dataset Format:** XES (eXtensible Event Stream)
 - **Analysis Tool:** Apromore
 - **Attribute Classification:**
 - **Case Attributes:** Application type, loan goal, requested amount.
 - **Event Attributes:** Activity, resource, time stamp, monthly cost, offered amount, ...

Dataset Analysis

- **Adapted dataset**
 - **Dataset Format:** Encoded in CSV
 - **Transformation:** Dataset underwent 4 transformation steps
 - **Key Attributes:**
 - Includes KPI-related attributes like **duration**
 - **treatment** attribute to determine if a case has been adapted

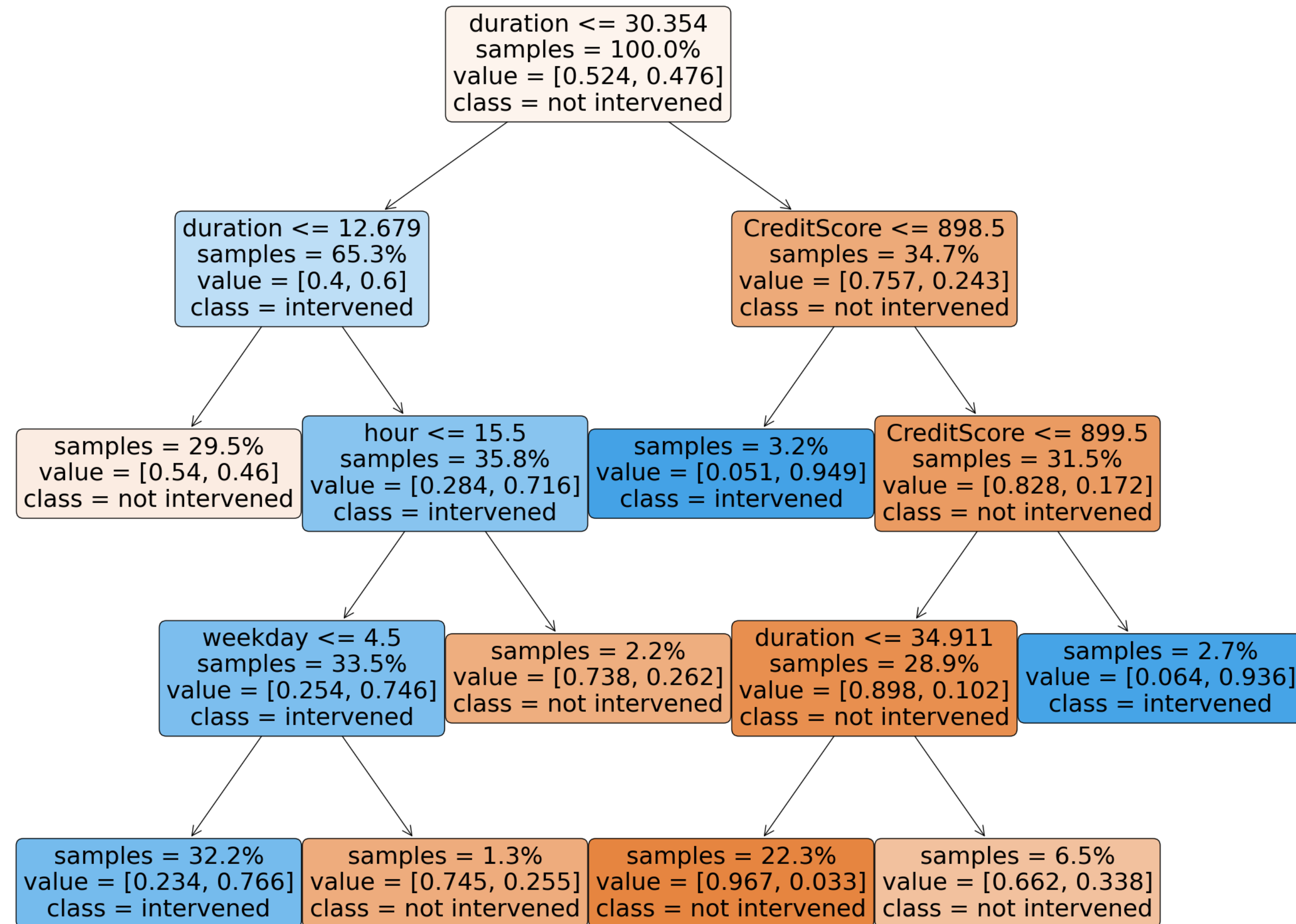
Implementation

Approach



Implementation

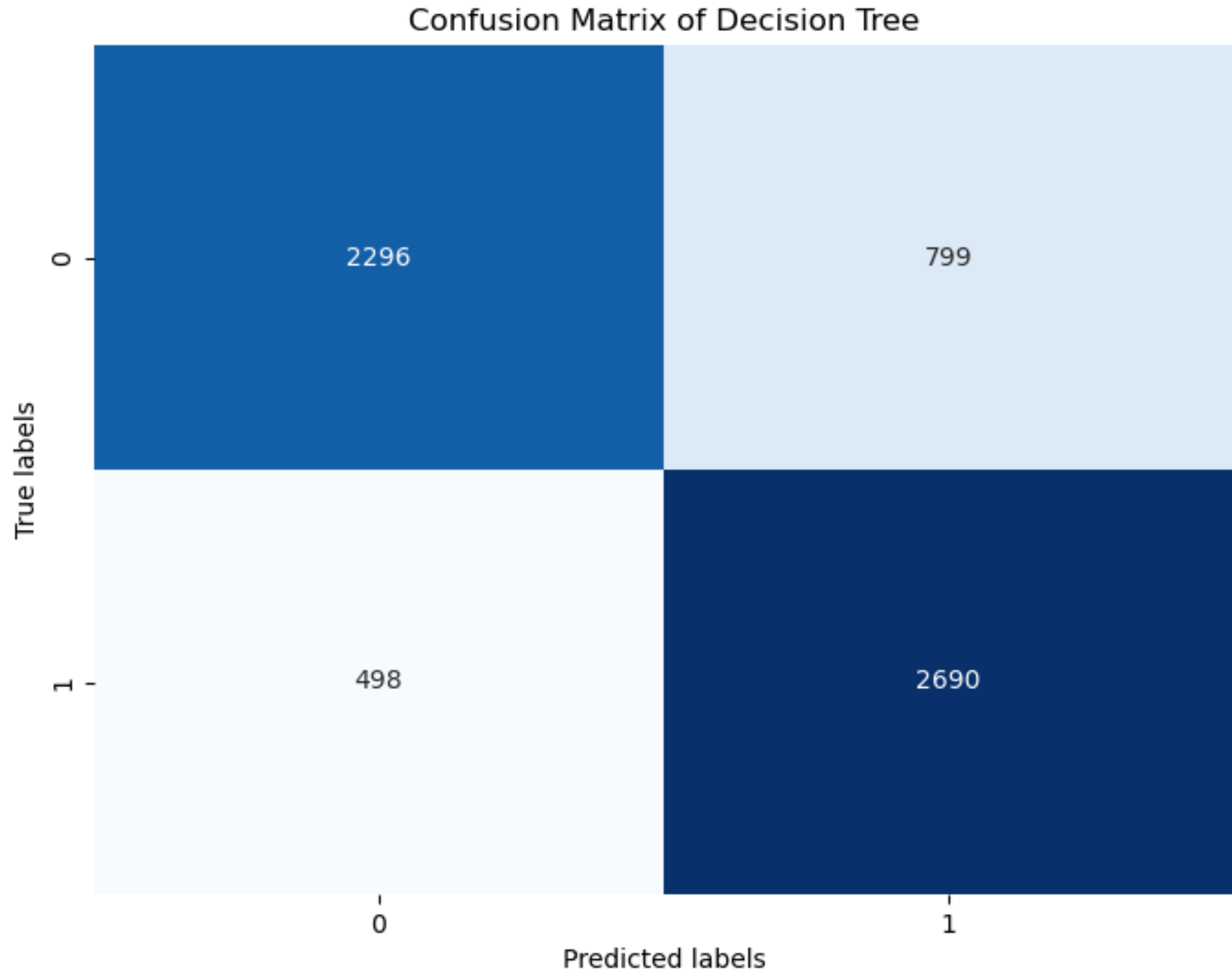
- **Decision Tree**
- **Key Attributes:**
 - Duration
 - CreditScore
 - Hour
 - Weekday



Results

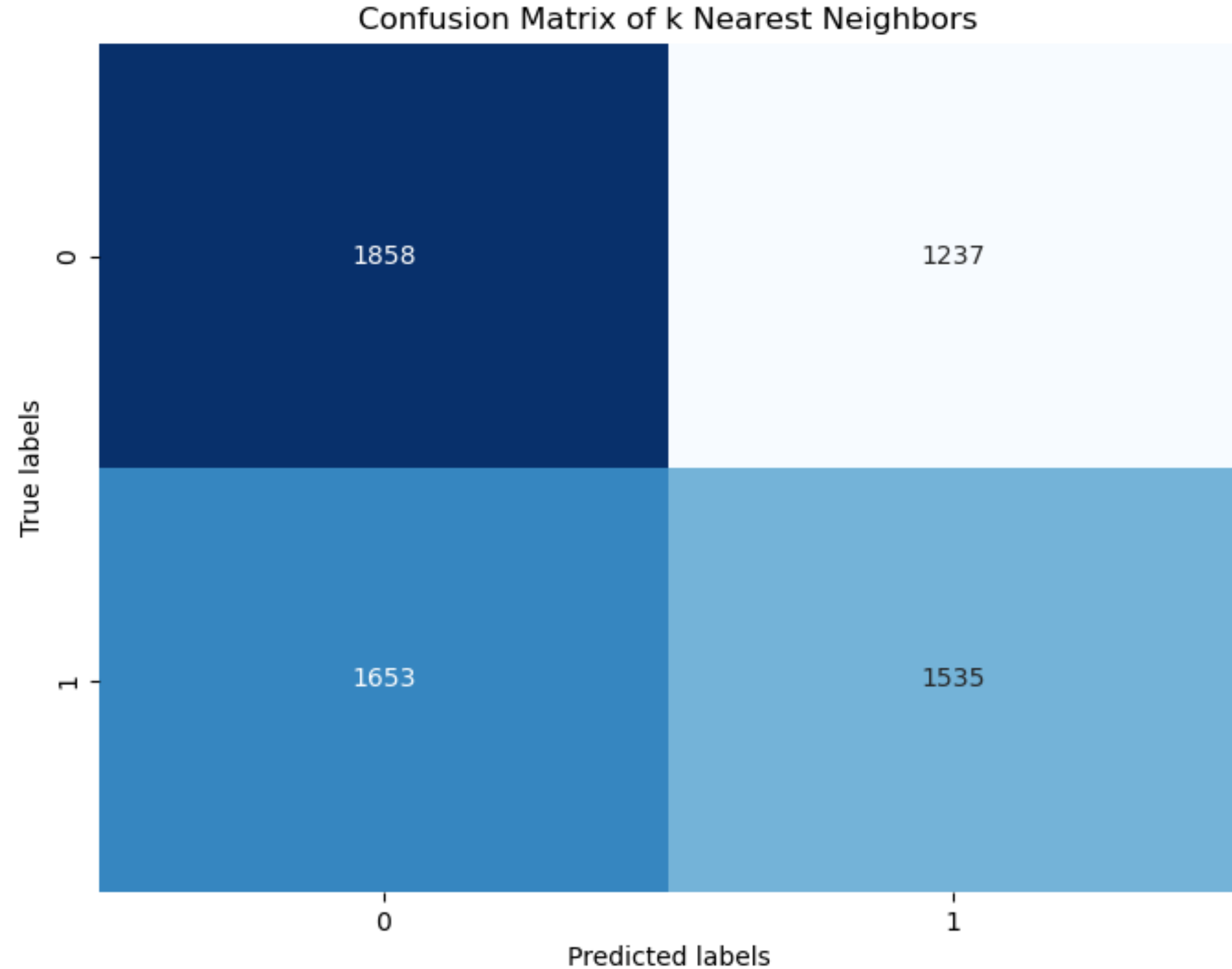
Results

- Confusion Matrix of Decision Tree



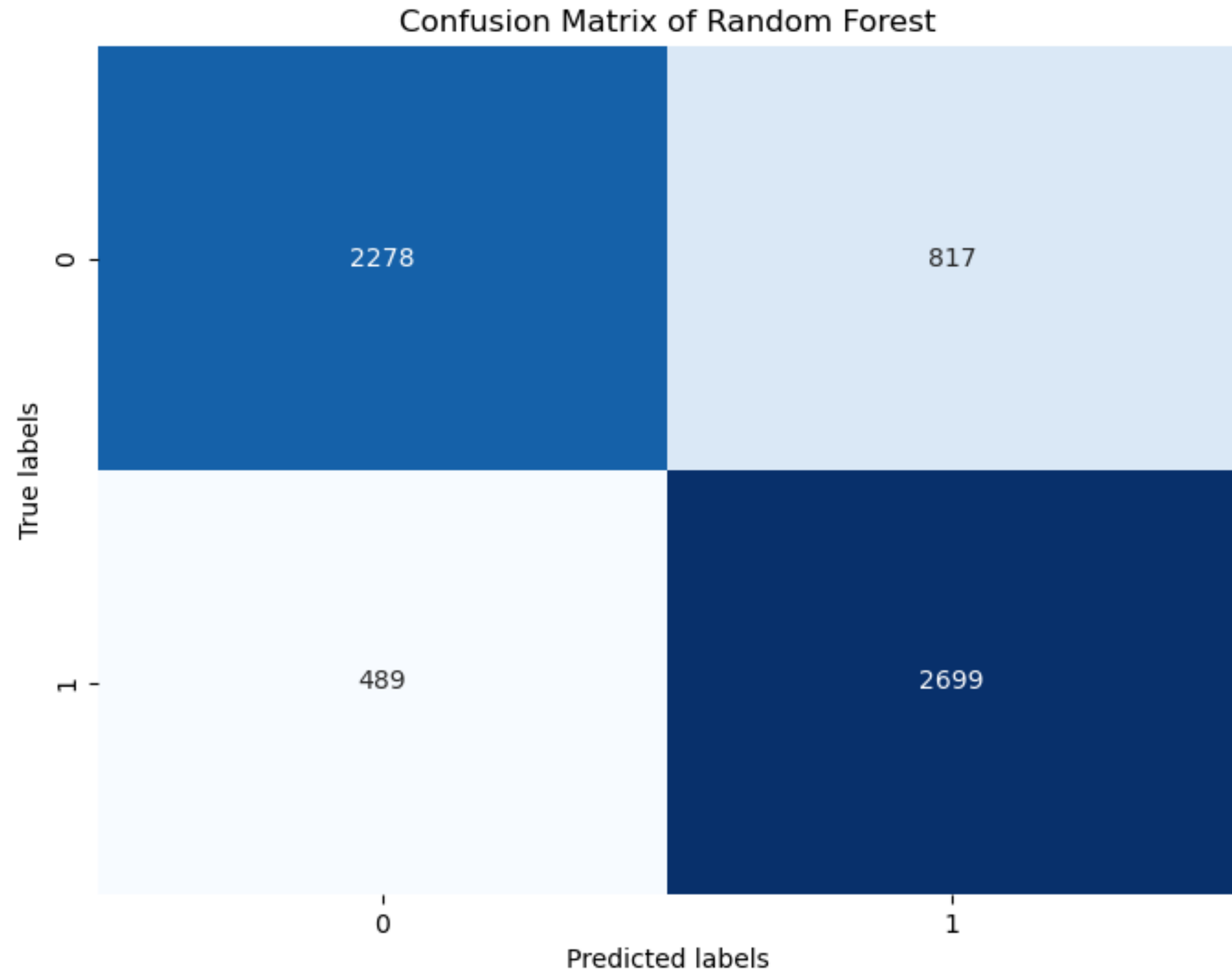
Results

- Confusion Matrix of k-Nearest Neighbors



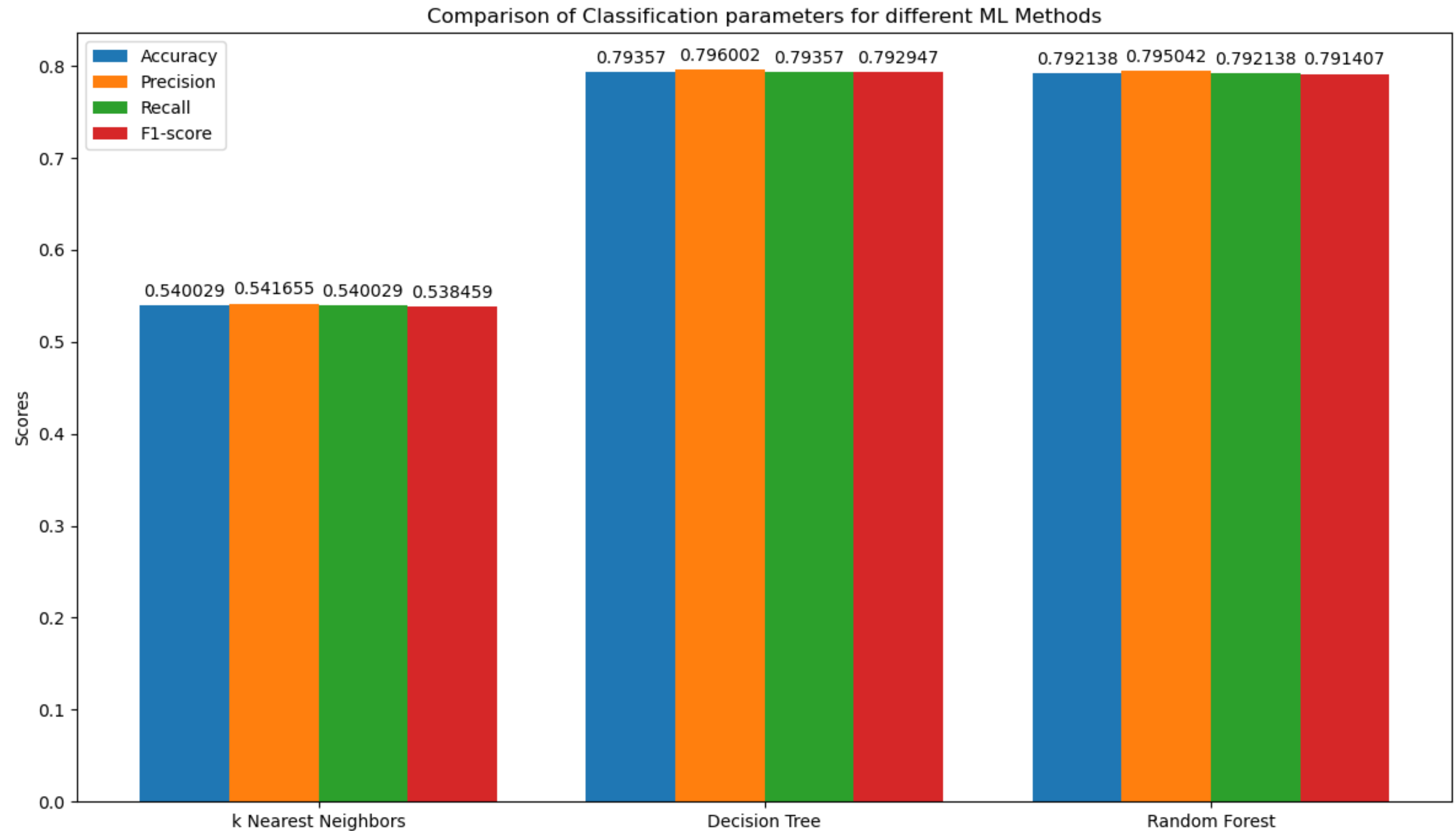
Results

- Confusion Matrix of Random Forest



Results

- Comparison



Conclusion

Conclusion and Future work

- **Objective**
 - **Difference between the Raw and Adapted Datasets**
 - **Approach to implement Change Detection**
 - **Correlation between Adaptation and KPI-attributes**
- **Limitation:** Scarcity of real datasets related to KPIs. Only one dataset was used.
- **Future Direction:** Investigate more datasets and explore other explainable machine learning algorithms.



Thank you for you attention!

References

1) Tools:

- 1) Logo of Jupyter Notebook: <https://jupyter.org/>
- 2) Logo of Scikit-Learn: <https://scikit-learn.org/stable/>
- 3) Logo of Apromore: <https://apromore.com/>

2) Datasets:

- 1) Raw dataset: https://data.4tu.nl/articles/dataset/BPI_Challenge_2017_-_Offer_log/12705737
- 2) Adapted dataset: <https://zenodo.org/record/5084612#.Y-JgtezMl-Q>
- 3) Source codes and outputs: <https://github.com/SiheonLee/bachelor-project>

