Project Overview: Predicting Customer Satisfaction with KNIME

Introduction

This project focuses on predicting customer satisfaction levels using a dataset from an airline company. We applied three different machine learning classifiers: Decision Tree, Random Forest, and K-Nearest Neighbour (KNN), utilizing the KNIME analytics platform to build and evaluate these models.

Objectives

- To predict customer satisfaction based on various service parameters.
- To compare the effectiveness of three classifiers in handling the prediction task.
- To determine the best model based on accuracy and other performance metrics.

Data Description

We utilized two main datasets:

- **Training Data**: Contains features and a target variable 'Satisfaction' indicating levels of customer satisfaction.
- **Test Data**: Used to evaluate the model by predicting the satisfaction level.

Methodology

Data Pre-processing

Steps included:

1. **Data Loading**: Utilizing CSV Reader nodes to load training and test data.
2. **Column Filtering**: Selecting relevant columns using the Column Filter node.
3. **Handling Missing Values**: Applying strategies to manage missing data.
4. **Data Normalization**: Standardizing data to improve model performance.

Model Building

- **Decision Tree**

- **Random Forest**
- **K-Nearest Neighbors (KNN)**

Each model followed a workflow of training with the pre-processed data, predicting using the test dataset, and evaluating performance through accuracy metrics.

Results and Findings

- **Decision Tree**: Achieved an accuracy of 99.105%, with a simple and interpretable model.
- **Random Forest**: Showed the highest accuracy at 99.996%, indicating robust performance against overfitting.
- **K-Nearest Neighbor**: Reported accuracy of 95.433%, effective for the dataset but slightly less accurate compared to other models.

Best Model

The **Random Forest** classifier was identified as the most effective model due to its high accuracy and robustness in handling the prediction task.

Conclusion

This analysis demonstrated the power of ensemble learning techniques in improving prediction accuracy and reliability. Random Forest, with its ensemble approach, provided the most reliable predictions for customer satisfaction.

Future Work

Further analysis could explore:

- Applying more advanced machine learning algorithms like Support Vector Machines or Neural Networks.
- Including more features that could influence customer satisfaction.
- Utilizing cross-validation techniques to enhance model reliability.

Acknowledgements

This project was developed using KNIME, an open-source data analytics platform that supports various machine learning algorithms and data pre-processing techniques.