Body performance

Mona Alrussaini , Hind Alharbi

Abstract:

This project aims to determine if the employee will get the promotion or not by using machine learning, according to the various variables and data of the HR department, which are related to the employee's education level, the employee's performance, the employee's years of service and other information related to the employee, the data set that was trained and tested in this project Provided by Kaggle and labeled using classification models, with Neural Network and get approximately 61% accuracy

Design:

This project is one of the T5 Data Science BootCamp requirements. Data provided by Kaggle has been used in this project. The HR Analytics Classification and the most important characteristics and data included: no_of_trainings', 'age', 'previous_year_rating', 'length_of_service', 'referred', 'sourcing' to classify the employee's attainment As promotion/not promotion.

Data:

The data set provided from Kaggle in .csv format, Contains 54,808 records,14 features each containing important metrics that can help categorize a person to be promoted or not. These features are cleaned up for ease of use and implementation using Ranking Modeling in ML.

Algorithm:

Features engineering:

- Checking for missing values and fill it .
- Rename features to easier and valid names to call and use .
- Identifying label features unique values.
- Cleaning label feature from spaces, lower cases if there is.
- Replace string values into numerical then change feature type to int.
- Selecting unique variable in label feature to be categorical features.

Models:

Random Forest, Support Vector Machine, were used before Neural Network model classifier that get the higher accuracy.

Model evaluation and selection:

The entire dataset containing 54,808 records, every record has 13 features splitinto 80/20 train vs test, and all scores reported below were calculated with classification report on both splits.

Final results of using Neural Network:

Accuracy: 61%

Precision: 62%

Recall: 61%

F1: 61%

Tools:

- Pandas for data manipulation
- Neural Network for modeling
- Matplotlib and Seaborn for plotting

Communication:

The slides are provided besides details are provided at this document of theproject. Feel free to any pull requests.