T5 Data Science BootCamp

HUMAN RESOURCES ANALYTICS

Done by:

Arwa Alolyani ,Ghaidaa althobaiti

Submit to: Mr.Ali El-kassas

ABSTRACT

A company, which is active in Big Data and Data Science, wants to hire data scientists among people who successfully pass some courses, which conduct, by the company. Many people signup for their training. Company wants to know which of these candidates are really wants to work for the company after training or looking for a new employment because it helps to reduce the cost and time as well as the quality of training or planning the courses and categorization of candidates. Information related to demographics, education, experience are in hands from candidates signup and enrollment.

DESIGN

This project is one of the T5 Data Science BootCamp requirements. Data provided By:

- HR Analytics: Job Change of Data Scientists.

DATA

We selected data HR Analytics: Job Change of Data Scientists containing more than 19000 rows , 18 columns and this data is imbalanced, Most features are categorical, so we Use the Resampling strategies and apply models to choose the best score of predict

ALGORITHMS

1.Download and read data
Divide the data into 3 sections: Training, Verification and Test
2. Quick Look at the Data Structure
Looking for Correlations and value count for every column
3. Prepare the Data
for Machine Learning Algorithms
Fill null values ,Replaces some values in some columns
4. Feature Engineering
Extract experience level based on experience column , Extract company size based on company column.
Converter object columns type to int46 .
convert all columns to dummy columns before start a classifier.
5.Models
We applied four different rating models
KNeighborsClassifier,LogisticRegressionCV,
DecisionTreeClassifier,RandomForestClassifier.
Their performance was evaluated

TOOLS

Library :
Pandas , Numpy,Math ,Matplotlib.pyplot,Sklearn ,Scipy,
Sklearn.linear_model ,Statsmodels.api and
make_classification,StratifiedKFold,accuracy_score,
imblearn.over_sampling sklearn.linear_model, sklearn.metrics
sklearn.tree.
Program : Python ,Jypter,Visme,Github