Tamader AboAlhassan - Afrah Almoutiri - Moneera Alfulaij

أكاديمية سدايا
SDAIA Academy

Resume Screening for Taqat using NLP

Problem

Large companies do not have enough time to open each resume, and they need a way to filter the resumes based on fields and positions. So, by using Natural language processing (NLP), topic modelling, classification, clustering and different techniques for the Resume Screening task.

Data Description

The data that will be used in this project has been taken from Kaggle website (https://www.kaggle.com). It includes data such as: (Category and Resume) It includes nearly 1000 rows.

Algorithms

After we loaded our data, we transformed it into data frame, we have done the cleaning and text preprocessor, and then we did EAD. After that we used libraries in TOOLS to do.

Tools

- Numpy and Pandas for data manipulation.
- Matplotlib and Seaborn for plotting.
- LogisticRegression model from sklearn.linear_model class to build a classification algorithm that is used to predict if the client will subscribe.
- train_test_split function in Sklearn model selection for splitting data.
- KNeighborsClassifier model from sklearn.neighbors to build a classification algorithm that is used to predict if the client will subscribe.
- DecisionTreeClassifier model from sklearn.tree to build a classification algorithm that is used to predict if the client will subscribe.
- RandomForestClassifier model from sklearn.ensemble to build a classification algorithm that is used to predict if the client will subscribe.

- Measure performance of each algorithm using precision_score, recall_score, accuracy_score, roc_auc_score and confusion_matrix from sklearn.metrics module.
- Show the report about the data using ProfileReport from pandas_profiling.
- Jupyter notebook to execute the code.

Communication

Presentation.