Employee Termination Prediction

T5 BOOTCAMP DATA SCIENCE PROJECT

HAYA JALAL

Project Goal

Predict the termination of employee, by analyzing the data using machine learning models.

Data

The dataset is publicly available in Kaggle, and is provided in .csv format. It contains 49,653 instances with 18 variables.

Data Preprocessing

- Drop unnecessary columns
- Drop all columns not available before termination
- Building Pipeline
- Train-test split (70/30)

Modelling

Logistic Regression model was used to predict the termination. The official metric was the accuracy of the model.

The accuracy of the model was:

- Accuracy: 100% "Perfect Model"

Predict the termination

In [19]:

```
sample_input = pd.DataFrame(pd.Series({
      'recorddate_key': '6/1/2006',
       'birthdate_key': '6/28/1944',
    'orighiredate_key': '1/3/1993',
                 'age': 65,
           'city_name': 'Fort St John',
     'department_name': 'Dairy',
           'job_title': 'Dairy Person',
          'store_name': 12,
         'gender_full': 'Female',
         'STATUS_YEAR': 2006,
       'BUSINESS_UNIT': 'STORES'
})).T

prediction = model.predict(sample_input)
print("Model Prediction:", prediction)
```

```
Model Prediction: ['TERMINATED']
```

Future Works

- Apply more models in the data set like random forest, naïve bayes, …etc
- Apply another metrics like precision, recall, roc curve and auc, …etc

THANK YOU!