The Data Science Course : 2020

Complete the Data Science Bootcamp

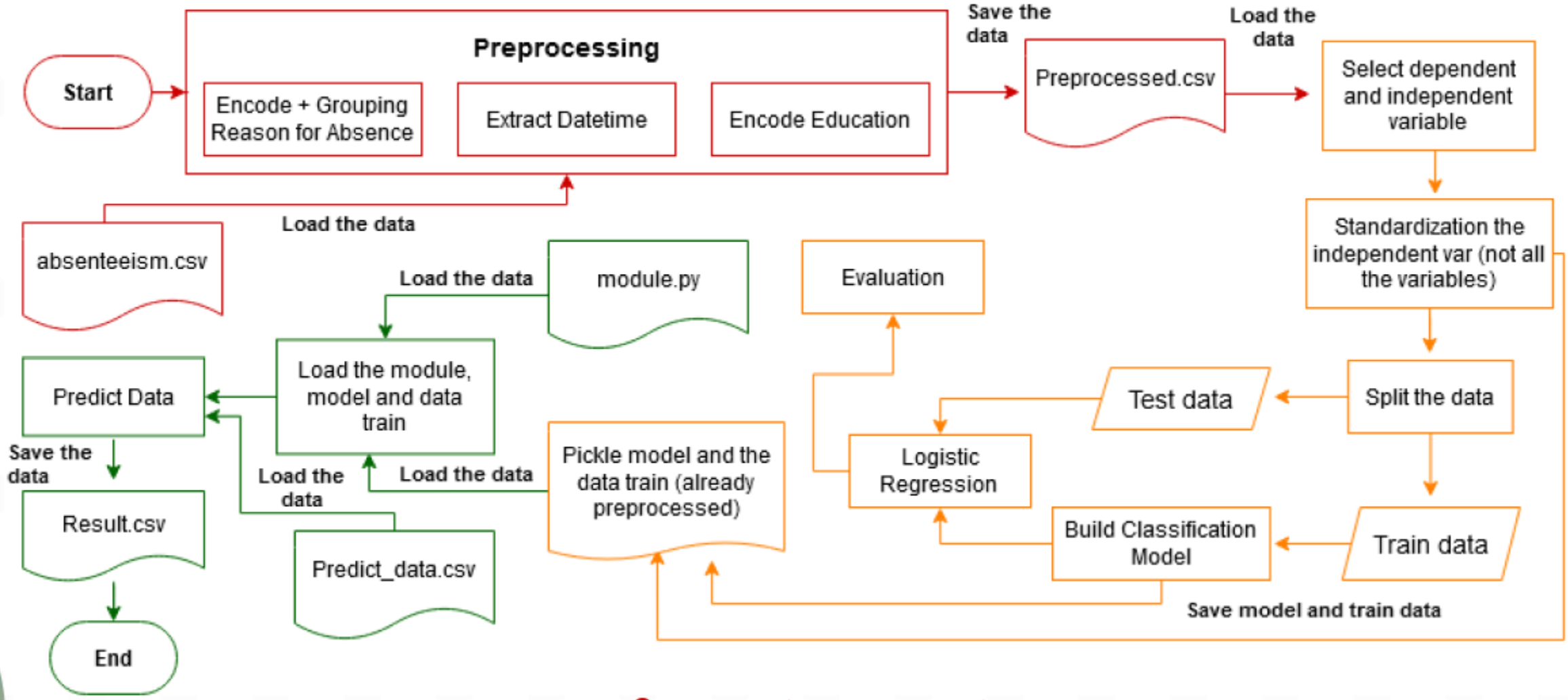
FREE COURSE
RESUME

Udemy

365 Careers

Predict Absenteeism

Using Logistic Regression (Last Project)



- Preprocessing & Data Extraction
- Data Modeling
- Model Integration

Notes For the Model Workflow Based on the Previous Slide

1 Divides the program file into three parts

The first file is used to analyze, extract and take some insight from the data (simple preprocessing).

The second file is used to build classification model, test the model and save train data + model using pickle.

The Third file is used to load our model and try to predict absenteeism value from data with empty label (model integration).

2 Data Preprocessing

I have done analyst the absenteeism reason and encoding the data in this process. But you can try to **re-encode the data** (using label or one hot encoding), **make the data normally distributed** (log, square root, box-cox, yeo-johson, etc), **scale the data** (StandardScaler), **imbalance class handling** (undersampling or over sampling), etc.

3 Build Model Classification

In this stape, I did not doing ***hyperparameter tuning*** to get the best parameter for the classifier (gridsearchcv or randomizedsearchcv). So you can try to implement this step if you run this code. Anyway, you can also ***try other classifer model***, then compere the model accuracy and chose the best one.

4 Save model and train data

The easiest way to save the model and the data train is using ***pickle library***.By save the model we don't need to re-train the classification model that wasted a lot of time. Just load the model in our module and start to predict new data.

5 Model Integration

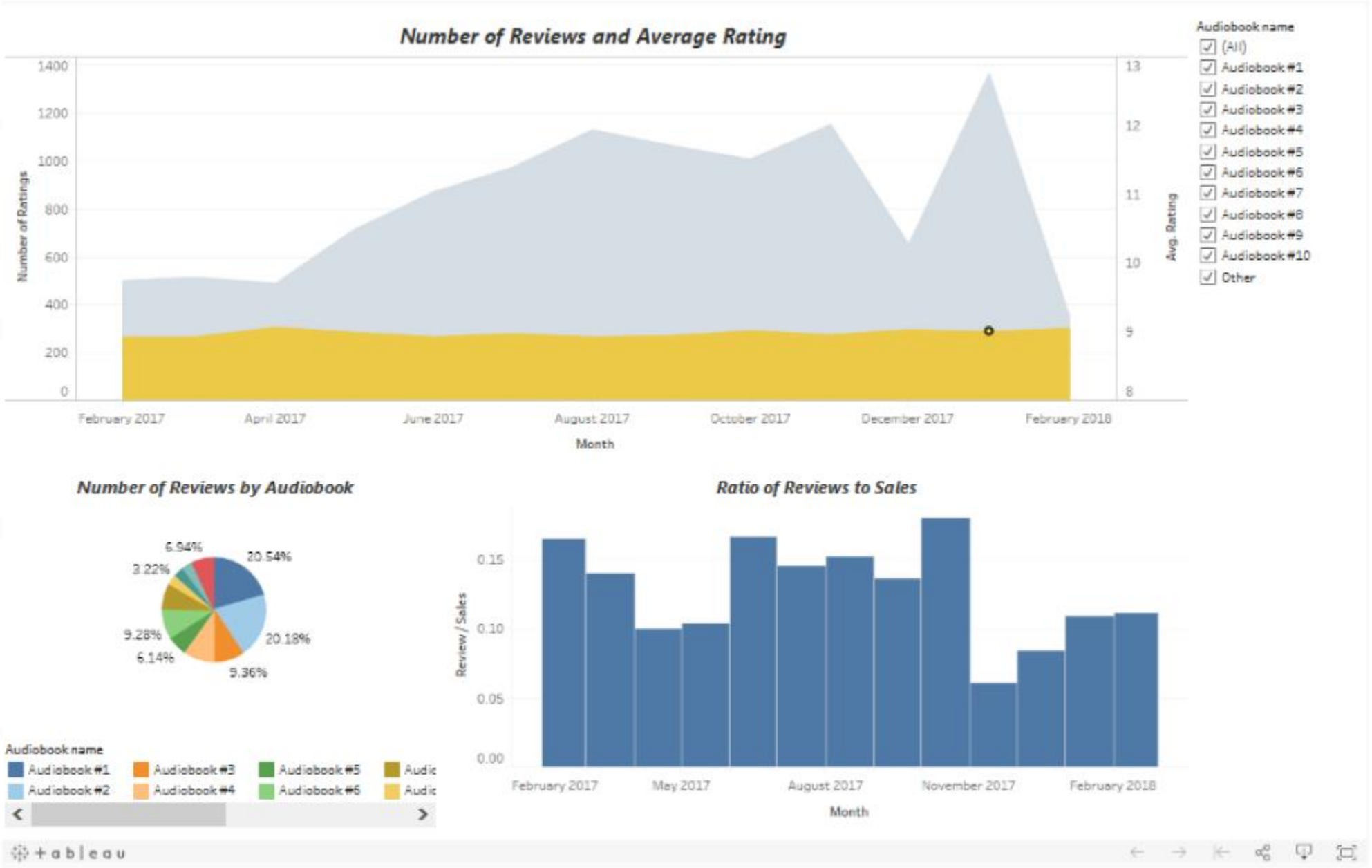
To integrate the model (step 4), you need to create a ***module*** (usually in .py format) where it can be use to load the model. This module is usually a class with various functions inside of it.

6 Predict the data

Firstly you need to import the module, use our model and train data inside the module then run the program. Now is the time to predict the new data by calling the module function.

Bonus Session, "analyze the result in Tableu Dasboard to get more insight from it."

This subject only focuses on model integration. So you can improve the model by yourself....



Visit my : [Tableu Account]

Certificate of Completion

This is to certify that ***Mohamad Irwan Afandi*** *successfully completed 29 total hours of* ***The Data Science Course 2020: Complete Data Science Bootcamp*** *online course on Sept. 8, 2020*

365 Careers
365 Careers, Instructor

365 Careers Team
365 Careers Team, Instructor

&

Udemy

Certificate no: UC-c40c9a3a-ae51-4325-b65a-851926526d88
Certificate url: ude.my/UC-c40c9a3a-ae51-4325-b65a-851926526d88

#BeAble

Start your journey

Download on my github account

irwanafandi24 Update README.md — `0389c92` yesterday — 3 commits

1. Simple Linear Regression Stats Model	machine learning basic	yesterday
10. Introduce to NN and Tensorflow	machine learning basic	yesterday
2. Multiple Linear Regression	machine learning basic	yesterday
3. Linear Regression with Prediction	machine learning basic	yesterday
4. Single Linear Regression SKlearn	machine learning basic	yesterday
5. Multiple Linear Regression SKlearn	machine learning basic	yesterday
6. Full Linear Regression	machine learning basic	yesterday
7. Basic Logistic Regression	machine learning basic	yesterday
8. Binary Logistic Regression Impleme...	machine learning basic	yesterday
9. Kmeans & Dendogram (Clustering)	machine learning basic	yesterday
Absenteeism Classification (Last Project)	machine learning basic	yesterday
README.md	Update README.md	yesterday