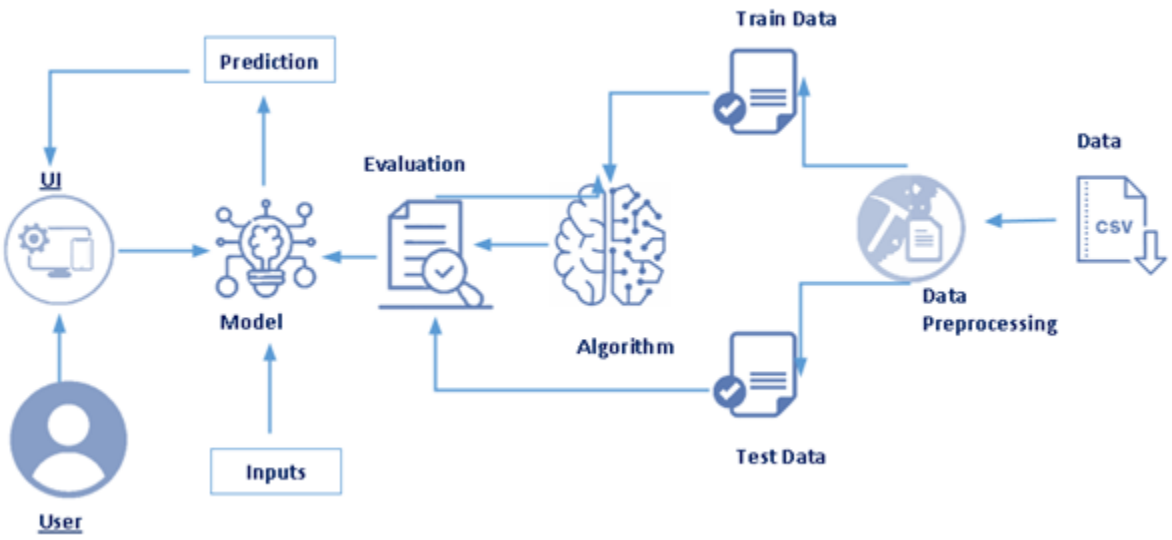
SESHADRI RAO GUDLAVALLERU ENGINEERING COLLEGE

Predicting Employee Attrition Using Random Forest With IBM Cloud



SUBMITTED BY:

MOHITH KOTHA

PIN:20485A0312

Content

1 INTRODUCTION

Abstract:

Now a day's Employee Attrition prediction become a major problem in the organizations. Employee Attrition is a big issue for the organizations specially when trained, technical and key employees leave for a better opportunity from the organization. This results in financial loss to replace a trained employee. Therefore, we use the current and past employee data to analyze the common reasons for employee attrition or attrition. For the prevention of employee attrition, we applied a well known classification methods, that is, Decision tree, Logistic Regression, SVM, KNN, Random Forest, Naive bayes methods on the human resource data. For this we implement feature selection method on the data and analysis the results to prevent employee attrition. This is helpful to companies to predict employee attrition, and also helpful to their economic growth by reducing their human resource cost.

Overview:

An employee would choose to join or depart an organization depending on many causes i.e. work environment, work place, gender equity, pay equity and many other. The rest of the employees may think about personal reasons for instance relocation due to family, maternity, health, issues with the managers or co-workers in a team. Employee attrition is a major problem for the organizations particularly when trained, technical and key employees leave for best opportunities from the organizations. This finally results into monetary loss to substitute a trained employee. Consequently, we utilize the present and past employee data to assess the familiar issues for employee attrition. The employee attrition identification helps in predicting and resolving the issues of attrition. We can use this data to stop the attrition rate of the employees. For this working we use some methodologies of data classification. Those methodologies are Decision

Tree (it is tree structure that comprises a branches, root node and leaf nodes. every internal node indicates a test on an attribute, every branch indicates the result of a test, and every leaf node holds a class label), Naive Bayes (it is a classification methodology depending on Bayes Theorem. A Navie Bayes classifier presumes that the existence of a specific in a class is unrelated to the existence of any other feature. For instance, a fruit may be measured to be an apple if it is red, round, and regarding 3 inches in diameter. Still if these features depend on each other or upon the presence of the rest of the features, all these properties autonomously contribute to the probability that this fruit is an apple) Logistic Regression(it is a statistical approach for assessing a dataset in which there are one or more autonomous variables that establish an outcome.

2 LITERATURE SURVEY

2.1 Existing problem:

Adding unstructured, textual data into a conventional attrition identification. The outcome is raise performance in attrition identification analysis. This study supportive for marketing decision makers to improved recognize customer those have probability to attrition.

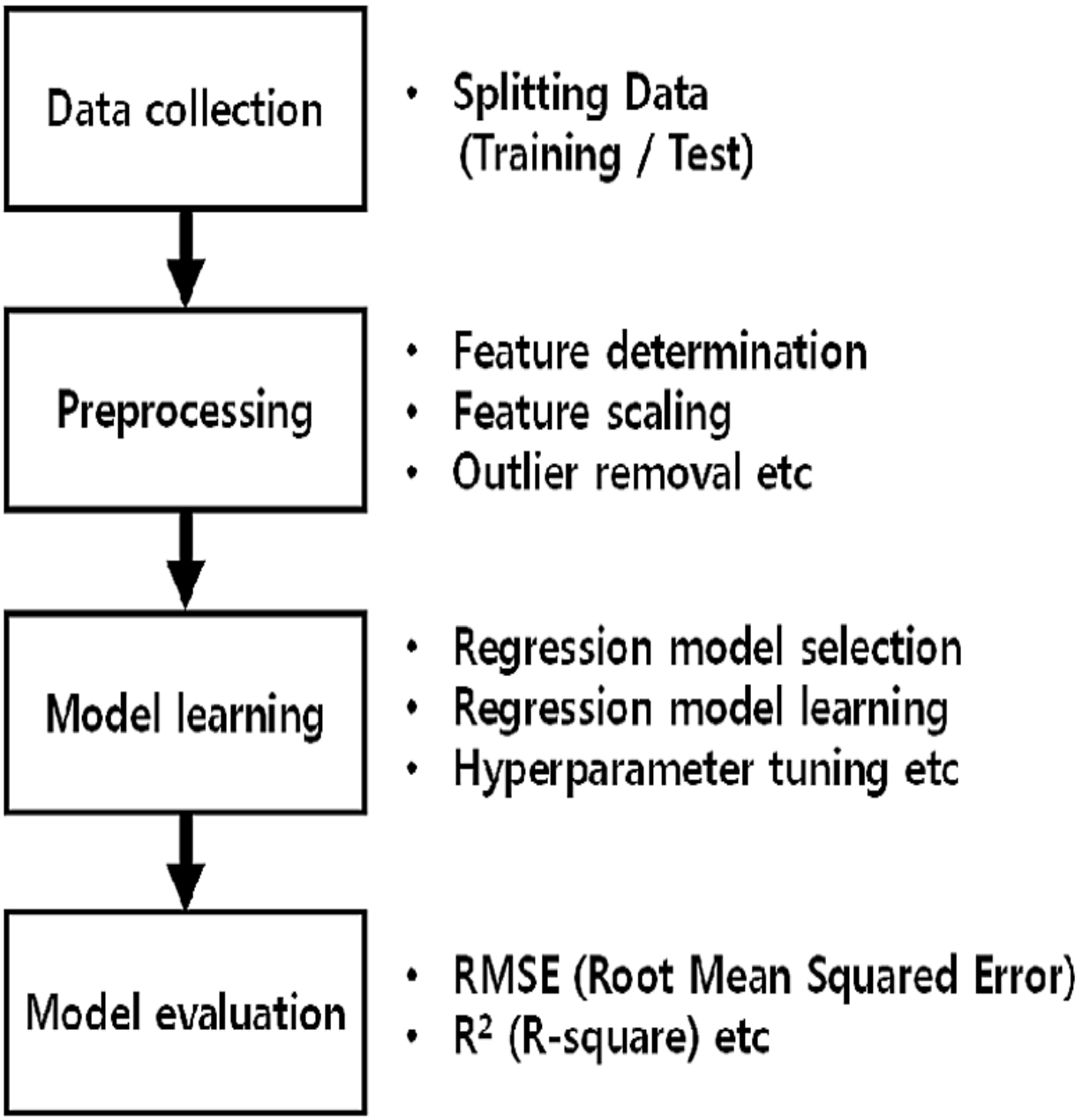
2.2 Proposed solution:

In the existing systems they used only few of data mining techniques for data prediction. Employee attrition effects in financial, time and effort loss for organizations. It is a big issue since a trained and experienced employee is difficult to substitute and it is cost effective.

so in this proposed method we use randaom forest regression for predicting emoployee atteration.

3 THEORITICAL ANALYSIS

3.1 BLOCK DIAGRAM:



3.2 Hardware / Software Requirements

Recommended System Requirements:

● Processors:Intel® Core™ i5 processor 4300M at 2.60 GHz or 2.59 GHz (1 socket, 2 cores, 2 threads per core), 8 GB of DRAMIntel® Xeon® processor E5-2698 v3 at 2.30 GHz (2 sockets, 16 cores each, 1 thread per core), 64 GB of DRAMIntel® Xeon Phi™ processor 7210 at 1.30 GHz (1 socket, 64 cores, 4 threads per core), 32 GB of DRAM, 16 GB of MCDRAM (flat mode enabled)

● Disk space: 2 to 3 GB

● Operating systems: Windows® 10, macOS*, and Linux*

Minimum System Requirements

● Processors: Intel Atom® processor or Intel® Core™ i3 processor

● Disk space: 1 GB

● Operating systems: Windows* 7 or later, macOS, and Linux

● Python* versions: 3.9

Software requirements:

anaconda navigator:

Anaconda is an open-source distribution for python and R. It is used for data science, machine learning, deep learning, etc. With the availability of more than 300 libraries for data science, it becomes fairly optimal for any programmer to work on anaconda for data science.

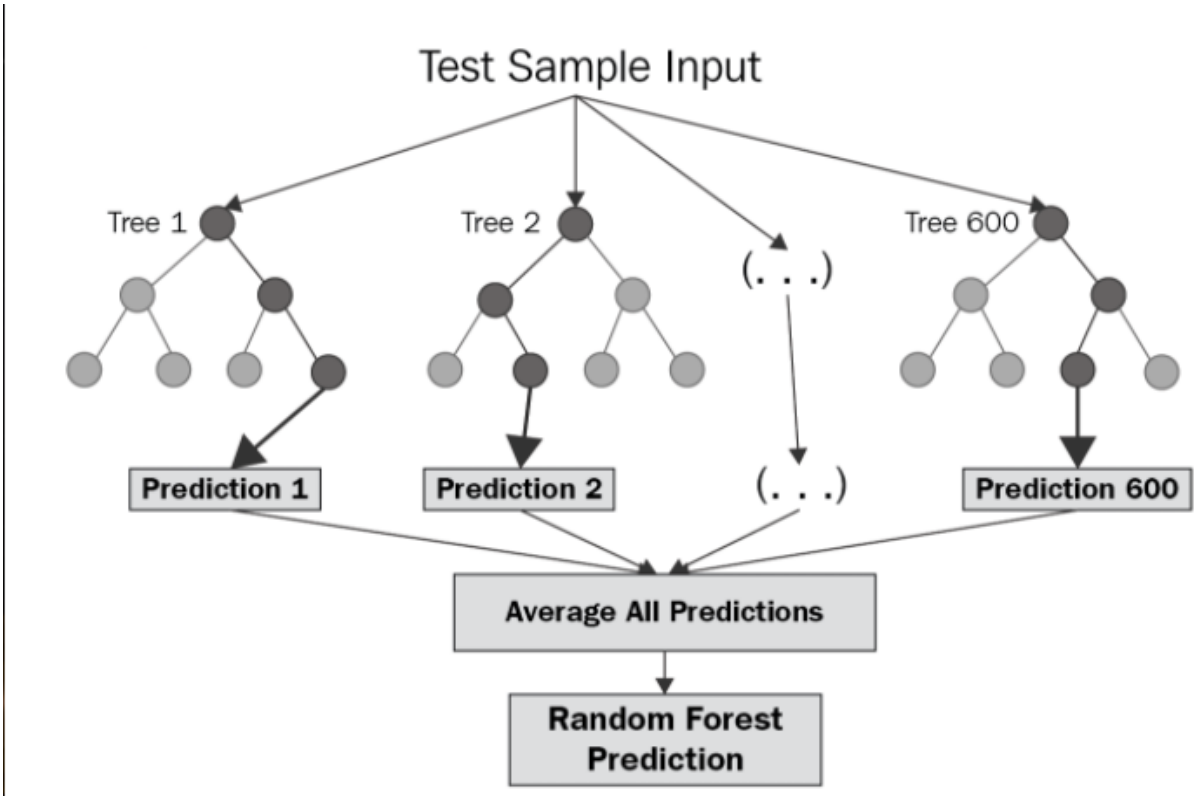
Pycharm:

PyCharm is a dedicated Python Integrated Development Environment (IDE) providing a wide range of essential tools for Python developers, tightly integrated to create a convenient environment for productive Python, web, and data science development.

4 EXPERIMENTAL INVESTIGATIONS

Random Forest Regression

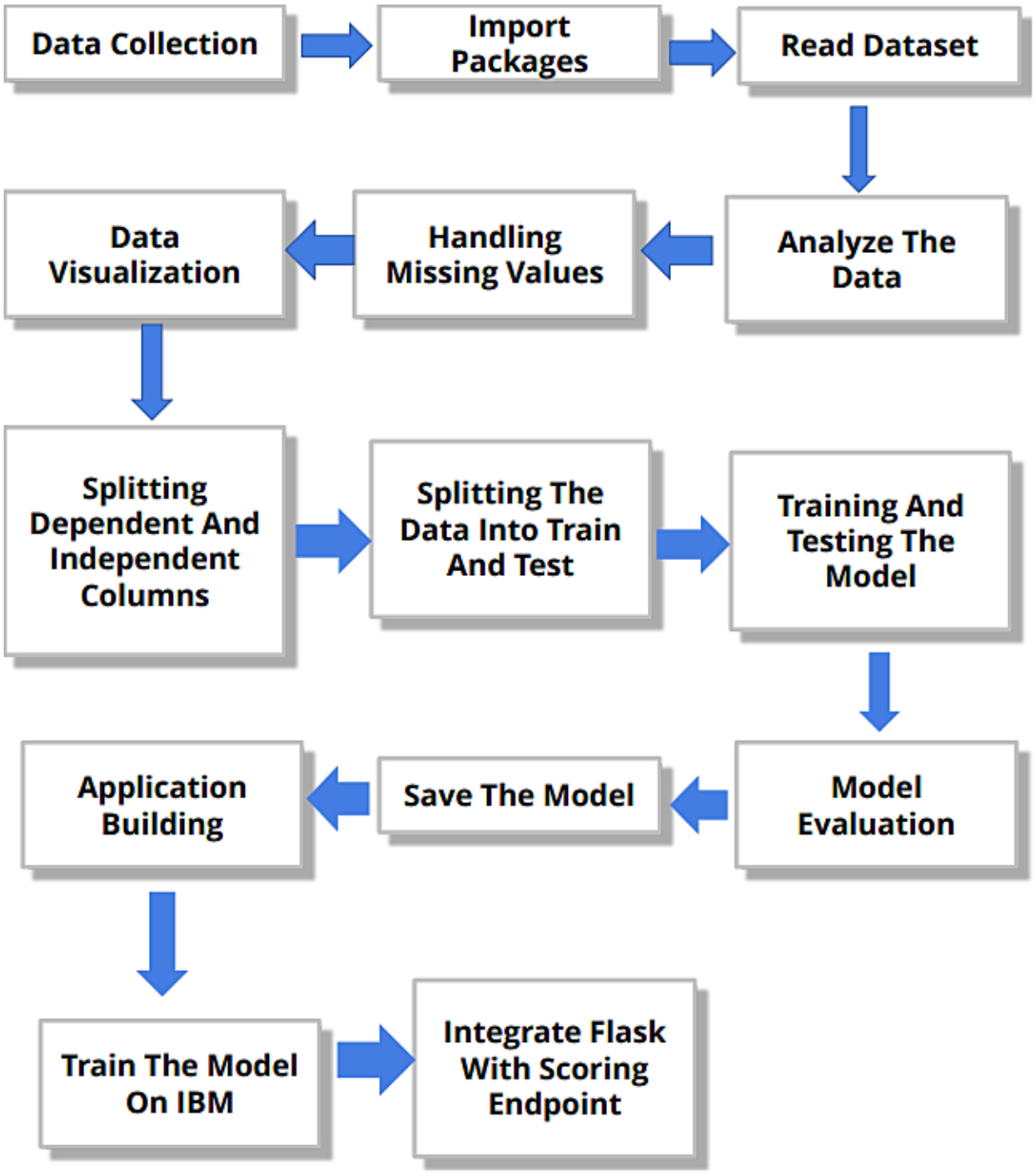
Random Forest Regression is a supervised learning algorithm that uses ensemble learning method for regression. Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model.



A Random Forest Regression model is powerful and accurate. It usually performs great on many problems, including features with non-linear relationships. Disadvantages, however, include the following: there is no interpretability, overfitting may easily occur, we

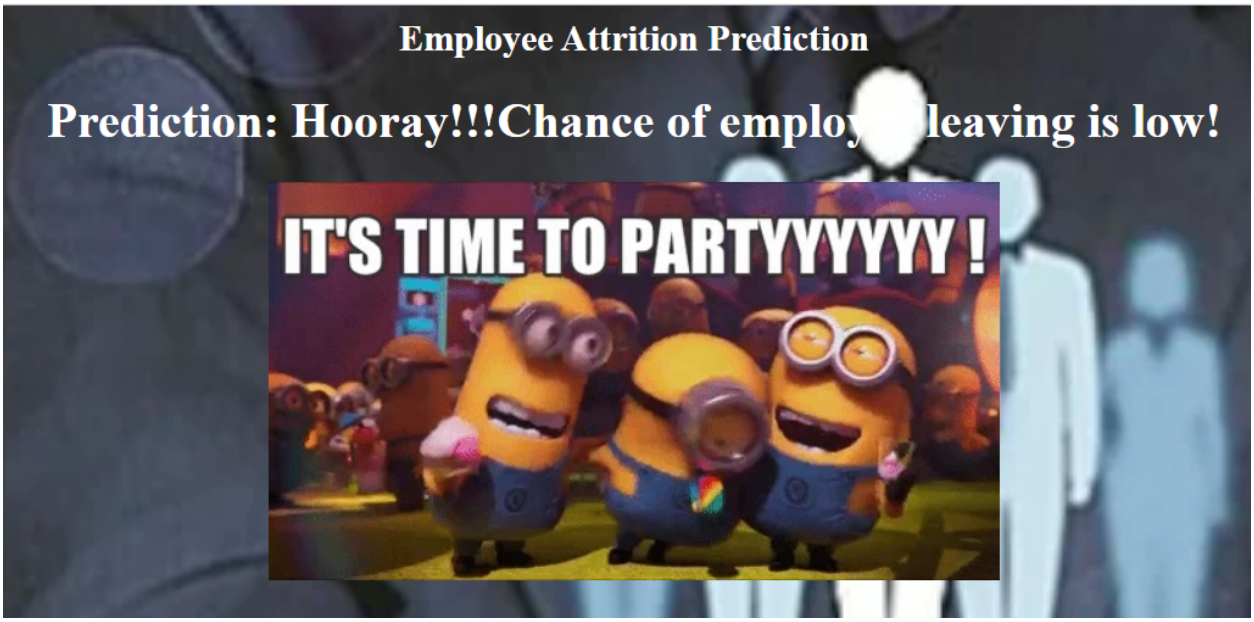
must choose the number of trees to include in the model.

5 FLOW CHART



6 RESULT

For knowing emoployee status. We will Enter Our Results Of All The Fields Enter Here And Check The Probability Here:



→If Student Getting Good Results Prediction Shows Your Eligible

→If Student Is Not Eligible It Shows You Are Chance of leaving is high Results

7 ADVANTAGES & DISADVANTAGES OF A PROJECT

We try to find to analyze the past and existing employee information to estimate the future attritions and study the reasons of employee turnover. The results of this learning describe that data extraction algorithms can be utilized to construct reliable and accurate predictive methods
for employee attration.

8 CONCLUSION

Employee attrition effects in financial, time and effort loss for organizations. It is a big issue since a trained and experienced employee is difficult to substitute and it is cost effective. We try to find to analyze the past and existing employee information to estimate the future attritioners and study the reasons of employee turnover. The results of this learning describe that data extraction algorithms can be utilized to construct reliable and accurate predictive methods for employee attrition. The issue of attrition identification is not just to depict attritioners from no attritioners. By using tentative data study and data extraction methids, we can depict the attrition probability for each one employee and provide them score to build the retention techniques.

9 future scope

Employee attrition is predictable under **stable circumstances**, wherein a set pattern can be deduced from certain parameters influencing the employee and the organization at all times. Some of these parameters could be foreseeable such as retirement age or unforeseeable such as company performance, external funding, management shakeup etc.

10 BIBILOGRAPHY

Installation of Anaconda Navigator:

https://www.youtube.com/embed/5mDYijMfSzs

Installation Of Pycharm Professionals:

https://www.youtube.com/embed/z73PyNDgVyQ

Installation Of Python Packages:
https://www.youtube.com/embed/akj3_wTploU
Data Collection:
https://www.kaggle.com/datasets/rishal005/admission-predict
Data Pre-processing:
https://thesmartbridge.com/documents/spsaimldocs/Datapreprocessing.pdf
Handling Null Values:
https://towardsdatascience.com/7-ways-to-handle-missing-values-in-machine-learning-1a6326adf79e
Data Visualization:
https://www.youtube.com/embed/TLdXM0A7SR8
Splitting Dependent And Independent Columns:
https://www.youtube.com/embed/A_V6daPQZIU
Splitting The Data Into Train And Test:
https://www.youtube.com/embed/xgDs0scjuuQ
Training And Testing The Model:
https://www.youtube.com/embed/yIYKR4sgzI8
Model Evaluation:
https://towardsdatascience.com/the-5-classification-evaluation-metrics-you-must-know-aa97784ff226
Flask Frame Work Referance:
https://www.youtube.com/embed/lj4I_CvBnt0
Flask Refarance To Run:
https://www.youtube.com/embed/UbCWoMf80PY
Train The Model On IBM:
Account Creation:
https://www.youtube.com/embed/4y_zD-0Q3F8
Train Model On IBM Watson&Integrate Flask With Scoring Endpoint
https://www.youtube.com/embed/TysuP3KgSzc
https://www.youtube.com/embed/ST1ZYLmYw2U

Thank You