AUTOMATIC DEPLOYMENT OF A.I MODEL

This MS Project report is submitted to the Department of Computer Science as partial fulfillment of Master of Science in Computer degree

by

KASHAF MANZOOR

Supervised by
DR TARIQ MAHMOOD
Professor
Department of Computer Science
School of Mathematics and Computer Science (SMCS)
Institute of Business Administration (IBA), Karachi

Fall 2021
Institute of Business Administration (IBA), Karachi, Pakistan

AUTOMATIC DEPLOYMENT OF A.I MODEL

This MS Project report is submitted to the Department of Computer Science as partial fulfillment of Master of Science in Computer degree

by

KASHAF MANZOOR
(ERP ID: 19836)

Supervisor:

DR Tariq Mahmood

Professor

School of Mathematics and Computer Science (SMCS)
Institute of Business Administration (IBA), Karachi

Fall 2021
Institute of Business Administration (IBA), Karachi, Pakistan

Dedication

I would be honored to dedicate this work to my parents who supported me throughout my life and allowed me to be what I am today. I also dedicate this work to my supervisor whose knowledge inspire me to work in this domain of A.I and last, all my teachers without their guidance I would not have been here.

Acknowledgement

I am Thankful to Allah and His blessings who allowed me to complete this project. I would like to express my special thanks of gratitude to my Supervisor Professor Dr Tariq Mahmood who gave me this opportunity to work on the trending technologies like Python, reactJS, firebase Realtime database. With his guidance, support and motivation play a very important role during my project work. At last, I would like to thank my parents for their continuous support which helped me a lot in finalizing this project.

Table of Contents

Abstract

Organizations where A.I models are deployed currently needs to retrain their model once accuracy goes down as the trends changes same model might not give the same result, so user usually train few models to see the result and deploy the best model. This cycle does not only cost time but money too and delays the process.

To reduce this overhead over organization we provided a solution where we train different models on data and calculate their accuracy the one with best accuracy will be deployed and make prediction for live data but once accuracy goes below defined threshold it will again goes for retraining models and again the one with best accuracy will be deployed.

This whole solution will only need to deploy once, and organizations can use this for a long time until they want to add a new model otherwise it will save a lot of time, money, and efforts.

Keywords: A.I model, retrain, redeploy, prediction, accuracy

1. Introduction

In recent years, companies are trying to understand the behavior of their customer/users to improve their business, so the best way is to implement A.I model which make the prediction for the companies therefore Companies invest into A.I model. Once they deployed their model, they continuously need to look over this model its accuracy and once result are not up to the requirement, they needed to retrain the model and redeploy it which is an expensive process and every time there is a possibility other models might perform better so its time taking process first analyze models than train it and deploy it.

Currently available products like Amazon AWS and Google cloud does not allow for retraining/redeployment of models they have wide variety of models from which user select model of their choice to train data and deploy it but once accuracy goes below requirement it again needs to be retrain manually and deployed again.

Our solution will make retraining/redeployment automatically we have currently added four models on which data is trained once data is trained on all four models, model accuracy is measured one with the highest accuracy is deployed and make predictions. So once it is deployed user is provided with a dashboard where user can look current model info and does not ever feel to retrain model manually it will retrain and redeploy automatically.

2. Background Knowledge

To understand this project there is a need to know some information before going into this project. User must know about dataset, a dataset is data of some company which have rows and columns, columns are features of data or we can say variables and rows are the values of variables. So, a clear understanding of dataset is required. Once dataset is provided, a model is trained, model is an A.I algorithm which utilize the dataset to produce some outcome this outcome is prediction of data. After a model is trained it is deployed, deploying a model means now model is ready to make prediction on live data.

3. Related Work

One of the available solutions is Google cloud machine learning engine which accepts a model and train data over it. It allows user to make changes in model and redeploy it so basically it takes a model and gives prediction over it. It will not retrain until user manually ask for it and it takes only one ML model for predictions [1].

Another available solution is AWS ML Services which gives user multiple choice to select for ML model it also automatically tweaks that model but to retrain (One click retrain) or to change the model user must do this manually [2].

The available solution provides retraining and redeployment of models but does not involve automatic redeployment. My solution will allow automatic retraining of model with capability of redeployment of A.I Model based on best performing model.

4. Problem Being Solved

A.I Model when deployed in industry it is trained over past data which is mainly comprise of few years of data. Once model is trained and deployed, it gives the industry accepted accuracy but over the time trends do change and this accuracy comes down and gradually deployed model starts to give unacceptable prediction. To overcome this model is retrained but with time better models are available in market but to change a model is an expensive move and usually company owners are not willing to invest that much.

We have identified two major problems:

1) Model needs to retrain over time once accuracy is below accepted threshold
2) Model needs to change if there is any other model which can produce better result

We are developing an A.I model where first it will train on given dataset. Once model is trained and prediction accuracy falls below 80% it will initiate a thread which will retrain this model and this newly trained model will be used and this process will be continue.

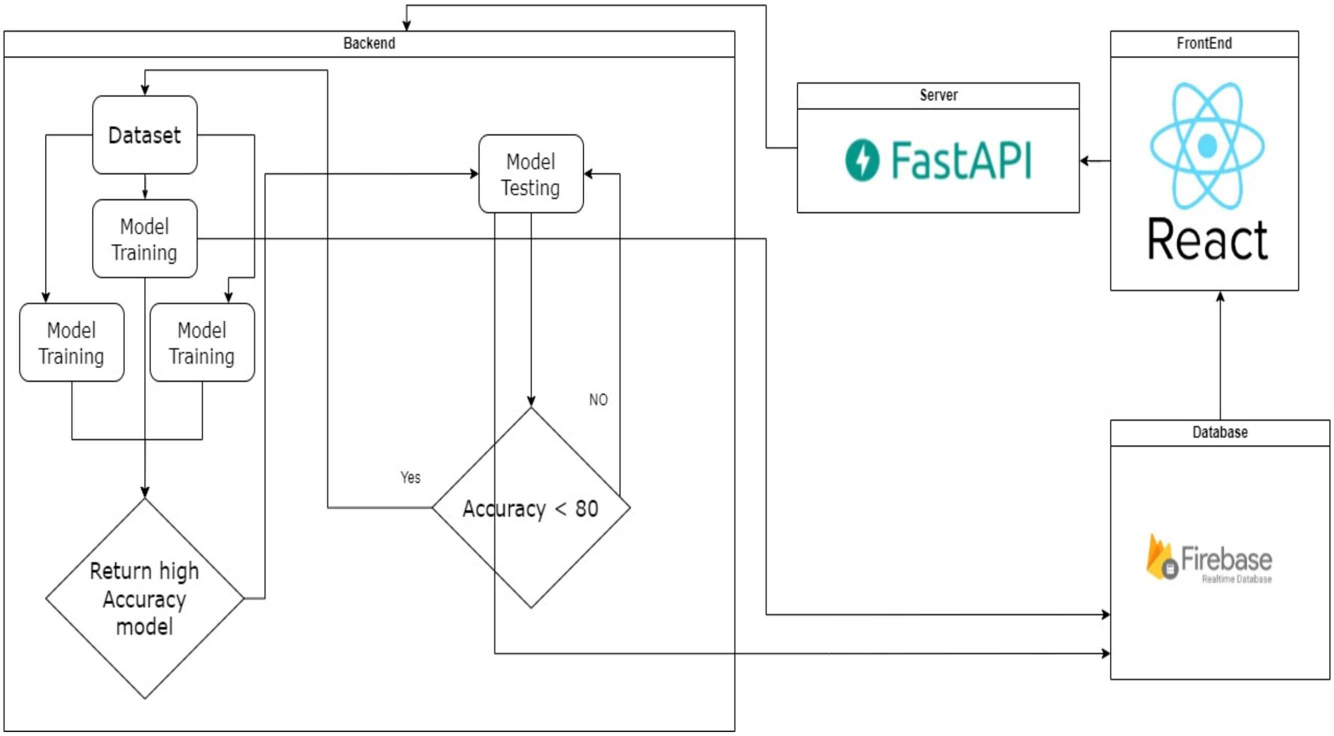
Training model will involve 4 different ML model the one which will have high training accuracy will be used.

ML Model which we will use:

1) Random Forest
2) Decision Tree
3) KNN
4) SVM (Support Vector Machine)

These models are selected because these are highly preferred models.

5. Methodology

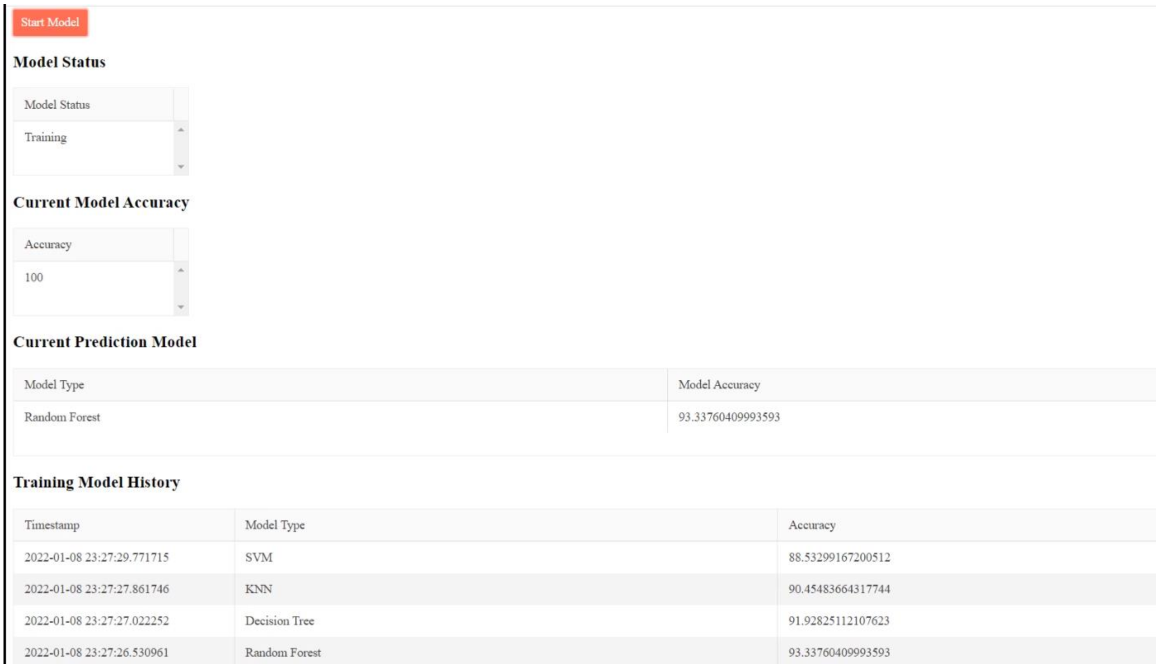


Our app is mainly made up with following elements

- Frontend is designed using ReactJS. ReactJS is JavaScript framework for creating and designing frontend where we are showing a dashboard for different statistics of models. React app is communicating with backend through API which is created on FastAPI. Data is coming from firebase real time database on dashboard.
- Backend is programmed in python programming language we train different data models on provided dataset and uses the one with high accuracy for prediction. While training/predicting we push some statistics to firebase and using firebase Realtime database and on frontend dashboard we are showing these statistics. if model accuracy goes below defined threshold which in picture can be seen as 80 than model will again goes into training and new model is trained using newer data and deployed. This will help us to overcome the changes in trends.
- FastAPI is python library which is used to create server where we have created an API which is called from frontend to start the model.
- Firebase Realtime database is used to display and change data on dashboard in real time. Data is being pushed into firebase database and data is displayed on frontend dashboard which is getting Realtime data update.
- Programing Languages used are JavaScript (ReactJS Framework) and python
- Programing software used to create this project are Nodejs, anaconda

- Following libraries are used before running this project
    - Sklearn is used for splitting data into test data and train data, and it is used for training models.
    - FastAPI is used to create API
    - Uvicorn is server environment it is needed to run FastAPI Server
    - Pandas is used for reading data from csv file and structure the data to be used in model

6. Software Application



Model Status

Model Status
Training

Current Model Accuracy

Accuracy
100

Current Prediction Model

Model Type	Model Accuracy
Random Forest	93.33760409993593

Training Model History

Timestamp	Model Type	Accuracy
2022-01-08 23:27:29.771715	SVM	88.53299167200512
2022-01-08 23:27:27.861746	KNN	90.45483664317744
2022-01-08 23:27:27.022252	Decision Tree	91.92825112107623
2022-01-08 23:27:26.530961	Random Forest	93.33760409993593

Green start model button is to start the application, once application is started this dashboard will show data like this.

- **Model Status:** This shows current state of model whether it is in Training status where models are training over dataset or Redeployed status where model is deployed and making prediction.
- **Current Model Accuracy:** This shows accuracy of deployed model it will goes down or up based on prediction it is making
- **Current Prediction Model:** This shows current Deployed model and model accuracy at time of deploying
- **Training Model History:** This shows Models and their accuracy in training

7. Conclusion & Future

Solution we have created will solve the problem of manually training a model and deploying it. Our solution will automatically deploy model and train dataset over four different model and the one with best accuracy will get deployed to make prediction and when its accuracy goes below defined threshold it will again train model with all the data it has made prediction and deploy new model.

Our solution solves the problem but there is still gap for improvement, handling for any dataset can be added, user input threshold and dynamically changing threshold. Currently this application has four model more models can be added and a solution for adding/removing models dynamically. A more customizable dashboard can also be provided to user.

References

1) Introduction to AI Platform | Google Cloud
2) Machine Learning | AWS AI and ML | Amazon Web Services (AWS)