Soumyadip Dutta

Data Scientist and Machine Learning Engineer

No 54, Gobi Prasanna Villa,
6th Cross, BDA Layout,
JeevanBima Nagar,
Bangalore-560017

+91 9903614919
soumyadipdutta2007@gmail.com

https://www.linkedin.com/in/soumyadip-dutta-7433041b

https://github.com/sdp1992

Being an engineer, I take pride in being a quick learner. Crisp deadlines and challenging tasks, make my day. I have traversed the field of electronics, moved on to the domain of data analytics and machine learning.

I expect to thrive in an organisation, which helps me building a steep learning curve, without compromising on quality and ability.

TECHNICAL SKILLS

Languages	Python, R, Scala
ML Algorithms	Regression, Tree Based Models, Boosting, Clustering, Time Series Analysis, Recommender System
Deep Learning Algorithms	Artificial Neural Network, CNN, Sequence Modelling, Natural Language Processing
Big Data Technologies	HDFS, Spark, Kafka, Sqoop, Hive
DL Frameworks	Tensorflow, Keras, PyTorch
Version Control Systems	GIT
Cloud Services	GCP, AWS, Azure
Devops Tool	Docker
Data Visualization Tools	Tableau, R Shiny
Others	SQL, NoSQL

EXPERIENCE

Harman Connected Services, Bengaluru — *Data Scientist*

September 2017 - February 2019

I am working as Data Scientist and ML Engineer. I do requirement gathering, client interaction and the required development.

ITC Infotech, Bengaluru — *Associate IT Consultant*

July 2015 - Aug 2017

I worked as Associate IT Consultant. I did requirement gathering, development and POC for clients.

EXPERIENCE

3 Year 7 Months

EDUCATION

Heritage Institute of Technology (WBUT), Kolkata
— B.Tech (ECE) 2011 - 2015
GPA - 7.73

Sarisha Ramakrishna Mission Sikshamandir (WBCHSE) *— Higher Secondary in Science 2011*

Sarisha Ramakrishna Mission Sikshamandir (WBBSE)
— Secondary 2009

PERSONAL PROJECTS

NaMo RaGa Twitter Battle, Github

Link- http://twitterbattle.reducetech.com

Technology - Tensorflow, Tensorflow serving, Spark Streaming,

Kafka, Flask, Sentimental analysis.

Tools: Google App Engine, Google Cloud, Digitalocean

Overview: This application projects last one hour tweet counts related to Narendra Modi and Rahul Gandhi.

And in Second part it shows sentiment wise count also(Count of positive and negative sentiment.)

I have collected tweets in real time from Twitter API and send it to Kafka topic. Then a spark streaming application is consuming those tweets from Topic.

Steps:

1. Counting tweets by person (Refreshing at 15 sec interval).
2. Making http request to tensorflow model server and getting sentiments of tweets.
3. Counting tweets by sentiment type as well as by person.(Refreshing at 15 sec interval)

COMPANY PROJECTS

Quote To Deal landing prediction, Harman Connected Services

September 2018 - Present

Technology - Python, Azure ML

Overview - Analyzing the risk of a quote. A quote may be or may not be landed as a confirmed deal. Analysing the risk of a quote based on more than 60 features.

My Role and Contribution -

- Created a risk analysis model based on XGBoost.
- Model tuning and continuous model maintenance.
- ⅛ th time reduction in number of manual risk analysis of quotes.
- Deployment in production.

PERSONAL DETAILS

Date of Birth	17th July, 1992
Nationality	Indian
Dialect	English, Bengali, Hindi

INTERESTS AND HOBBIES

IoT, Stock Market Trading, Cloud Computing

Targeted Marketing Campaign, Harman Connected Services

August 2018 - September 2018

Technology - Python, H2O

Overview -Increasing model performance of an existing model. This model predicts which customers are inclined to which genres of games. The false negative rates has been reduced drastically.

My Role and Contribution -

- Created a new challenger model using GBM.
- Tuned old hyperparameters.
- Reduced imbalanced class problem with proper sampling method.

Telecom Churn Product Development, Harman Connected Services

May 2018 - July 2018

Technology - Spark, R, R Shiny, Apache Livy Client

Overview - The application was built for internal purpose. Any customer can dump their telecom data and get predictions.

Custom hyperparameters and missing value imputation method can be selected from dropdown. Spark has been used for ML pipeline and R/R. Shiny has been used for visualization.

My Role and Contribution-

- Created ML Pipeline Spark.
- Commissioning HD Insight Cluster in Azure.
- Created R Shiny dashboard.
- The model has been created based on GBT classifier.

Event Sequence Analysis– BT, Harman Connected Services

April 2018 - April 2018

Technology -R, R Shiny, bupaR, edeaR, processmonitR

Overview -Created an event sequence flow based on call centre tickets. Created a optimized sequence path so that tickets can be resolved with minimum time. Created a predictive model to estimate the time and importance of the ticket.

My Role and Contribution -

- Created the event sequence timeline using R.
- Created a dashboard in Shiny to monitor event wise dwell time.
- Created an ML Model to predict time required to resolve a ticket based on importance.

Customer Convergence– BT, Harman Connected Services

September 2017 - March 2018

Technology - R, Python
Overview - This project aims at how two telecom companies can be merged retaining most of the customers. Clustering based on customer usage data and demographics data. Creating new products and offering those to customers so that customer attrition can be checked. Created a churn model to detect the customers who are most likely to churn.

My Role and Contribution :

- Detailed analysis how existing customer will behave in new company after the merging.
- Creating customized offers for customer to check churn.
- Created a ML model in XGBoost to predict potential churners.

Attrition Rate Analysis– ITC Infotech

October 2016 - July 2017

Technology - R

Overview - This ML model predicts which employees are more likely to leave the organization. 17 different features has been used for modelling.

My Role and Contribution -

- Using Gradient Boosting predicting the A.R of employees.
- Complete data cleaning and missing value imputation.

Attendance Pattern Analysis– ITC Infotech

January 2016 - October 2016

Technology - R

Overview - Attendance analysis pattern of employees across different business units. Finding anomaly in pattern.

My Role and Contribution -

- Worked as data analyst.
- Implemented predictive modelling using Linear Regression.
- Model maintenance and retraining.