VINAY SAMMANGI

@ vsammangi3@gatech.edu (+1) 770-703-0570 in VinaySammangi Research

EXPERIENCE

DATA SCIENCE INTERN @ Samsung
May'22 - Aug'22

- Developed hierarchical, multivariate time series models to forecast the hyperscalers demand using financial data and spending patterns.
- Deployed the models into a Streamlit application to forecast the demand for the next four quarters.
- Evaluating the current business forecasting process to ensure the ML model consistently outperforms the qualitative forecasts.

DATA SCIENTIST @ Aditya Birla Group
Apr'18 - Jul'21

Commodity Price Forecasting ($4M)

- Led a team in engineering highly relevant features affecting the prices using web-crawling, fundamental and technical analysis
- Forecasted prices with only **1% MAPE** using univariate and multivariate regression models that could work in uncertain market conditions
- Built robust classification models to predict the price movements with **100% accuracy** that provided confidence to the price forecasts

Industry 4.0 ($1M)

- Deployed a meta-learning model to predict the temperature spikes in a smelting process and provide recommendations for process control
- Solved quality improvement problems by identifying the root causes leading to process failure using anomaly detection methods

DATA ANALYST @ Meru & Ola
Jul'17 - Mar'18

- Identified the behavior of each customer using clustering analysis
- Predicted expected job time of a driver using ensemble models
- Implemented A/B testing to find the ideal mode of communication to the customer for either discount or dynamic fares

DATA SCIENCE INTERN @ Innoplexus
May'16 - Jul'16

- Used Web scraping techniques to automate the ETL processes by extracting **15 TB** of data from 60 different pharmaceutical websites
- Programmed a python class for parsing different file formats into JSON format and dumping them into MongoDB database

RESEARCH

Prediction and Prevention of Accidents
Nov'15 - May'17

- Evaluated different methods for handling missing values, outliers, generating new features, determining important features
- Identified hidden semantic structures of text using LDA topic modeling
- Worked extensively on hyperparameter tuning of SVM, ANN, Decision trees, Random Forests using PSO and GA algorithms
- Compared the model performances using statistical hypothesis tests
- Proposed a novel approach to extract rules from SVM and RF models
- Obtained predictive regions of large text data using CNN and developed the safety measures using Association Rule Mining
- Developed an android application that uses a BPNN model in the app to predict real-time health rate
- Collected the dynamic data from worker (heart rate) with the help of ICT based data capture system into the android application
- Published ***three journal papers, three conference papers*** from this project

EDUCATION

GEORGIA INSTITUTE OF TECHNOLOGY
Master of Science in Analytics

Aug'21 - Dec'22 (exp) Atlanta

CGPA: 4.0 | **Track:** Computational Data Analytics
Graduate Teaching Assistant:
Natural Language Processing, Machine Learning

INDIAN INSTITUTE OF TECHNOLOGY
B.Tech in Mechanical Engineering

Jul'13 - May'17 Kharagpur

SKILLS

Python | R | SQL | Flask | Streamlit | C | Tableau | NoSQL | Java | HTML | CSS | JavaScript | Azure ML | AWS | GCP | D3.js | PySpark | OpenRefine | SPSS | RapidMiner | Web Scraping | Optimization

COURSEWORK

Natural Language Processing | Deep Learning | Machine Learning | Regression Analysis | Data & Visual Analytics | Graphical Models in Machine Learning | Time Series Analysis | Data Analytics in Business | Analytical Modeling | Computing for Data Analysis | Business Fundamentals for Analytics

PROJECTS

Stock Price Forecasting Application:

- Predicted the Buy or Short signals by combining sentiment analysis and technical analysis on real-time price movements and tweets
- Developed an innovative multi-faceted web application with a simple interface to encourage retail investors to utilize a more quantitative approach to their trading strategy

Natural Language Processing:

- Developed a BiLSTM-CNN-CRF model for named entity recognition problem with 90.5% F1 score
- Implemented an encoder-decoder architecture with attention mechanism on dialog corpus
- Built an ensemble model comprising DistilBERT and Ridge to predict the degree of toxicity of social media comments with 85.3% accuracy
- Implemented a Convolutional Neural Network model on IMDB reviews with 87.2% accuracy.
- Developed a content-based recommendation engine on Netflix movies and TV shows data