



Vinay Sammangi

MS COMPUTATIONAL DATA ANALYTICS CANDIDATE AT GEORGIA TECH

I have the legal permit to work in United States

☎ (+1) 7707030570 | ✉ vsammangi3@gatech.edu | 🏠 Website | 🔗 LinkedIn | 🎓 Research

Summary

Innovative and result-oriented data scientist with 6+ years of research and professional experience collaborating with cross-functional teams, ensuring accuracy and integrity around data and actionable insights. Experience working with startups and conglomerates in different industries, providing solutions to price and demand forecasting, optimization, and root cause analysis problems, among many others. Published peer-reviewed machine learning research papers at international conferences and journals.

Education

GEORGIA INSTITUTE OF TECHNOLOGY

MASTER OF SCIENCE IN ANALYTICS

Atlanta, United States

Aug. 2021 - Dec. 2022 (exp)

- GPA: **4.0** out of 4.0, Track: Computational Data Analytics
- Graduate Teaching Assistant: Natural Language Processing (CS 7650 - Spring 2022, CS 4650 - Fall 2022)
- **First place** in 5 out of 6 data science projects among 120 students

INDIAN INSTITUTE OF TECHNOLOGY, KHARAGPUR

BTECH IN MECHANICAL ENGINEERING

Kharagpur, India

Jul. 2013 - May. 2017

Experience

NASA LANGLEY RESEARCH CENTER

DATA SCIENCE PRACTICUM

Remote, United States

Aug. 2022 - Dec. 2022 (exp)

Reconstitute dying and damaged coral reefs

- Develop python scripts to merge data from public reef databases into a dataset appropriate for machine learning.
- Time-Align and Geo-Align the coral reef dataset with the corresponding satellite imagery data.
- Use neural network models (based on satellite imagery) to demonstrate that we can predict the existence of coral in an observed region.

SAMSUNG

DATA SCIENCE ENGINEER INTERN

San Jose, United States

May. 2022 - Aug. 2022

Demand Forecasting for Hyperscalers

- Developed hierarchical, multivariate time series models using AutoML and HTS frameworks to forecast the quarterly hyperscalers demand by leveraging financial data and customer spending patterns.
- The forecasting **model outperformed the business forecasts** (Voice of Customer, Voice of Market) for the high-demand products. These products contribute to 80% of the overall demand.
- Deployed the solution and models into a streamlit application to forecast demand for the next six quarters.
- Detected product transitions using google trends data and analyzing market research reports.

ADITYA BIRLA GROUP

DATA SCIENTIST

Bangalore, India

Apr. 2018 - Jul. 2021

Commodities Price Forecasting (Traded and Non-traded) - \$4M

- **Supervised junior team members** in engineering relevant features affecting prices using web-crawling, fundamental and technical analysis.
- Forecasted prices with only **1% MAPE** using univariate and multivariate regression models that **worked in uncertain market conditions**.
- Built robust classification models to predict price movements with **100% accuracy**, thus providing confidence while using price forecasts.
- Productionized and monitored the models over two years, deployed the solutions using plumber API in R, and streamlit in Python.
- Identified and evaluated external data providers for macroeconomic factors and detailed supply-demand data to enhance the model's accuracy.

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.
- Predicted bad quality products with very high precision using classification techniques to improve the customer contentment.

Other Projects/ Achievements

- **Network Optimization:** Integer programming to reallocate warehouses and identify new locations using plant and customer data.
- Delivered **talks about short-time series forecasting** at IGIDR, B.K. Birla universities in India.
- Won internal hackathons focusing on Natural Language Processing and Computer Vision; Analyst of the Quarter in Q3 2020.
- **Collaborated with senior stakeholders** across different verticals (Hindalco, Thai Acrylic Fiber, Novelis, UltraTech).

MERU & OLA

DATA ANALYST

Mumbai, India
Jul. 2017 - Mar. 2018

- Performed customer-centric analysis to identify the behavior of each customer using clustering methods.
- 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.
- Extracted features from the driving license using Image Processing techniques.
- BI analytics gathering & reporting - developed python scripts to **automate day-to-day mundane tasks**.

INNOPLEXUS

DATA SCIENCE INTERN

Pune, India
May. 2016 - Jul. 2016

- Utilized web crawling techniques to automate the ETL processes; **scraped 15 TB of data** from 60 different pharmaceutical websites using beautiful soup and selenium frameworks.
- Programmed a python class for parsing different file formats into JSON format and dumped them into MongoDB.
- Created a binary document classifier on the IMDB and PubMed datasets.

TATA GROUP

RESEARCH ASSISTANT

Kharagpur, India
Dec. 2015 - May. 2017

Prevention of Accidents using Machine Learning

- Evaluated different methods for handling missing values and outliers, generating new features, and determining important features.
- Identified hidden semantic structures of text using LDA topic modeling.
- Worked extensively on hyperparameter tuning of SVM, ANN, Decision trees, and Random Forests using PSO and Genetic algorithms.
- Compared and contrasted the model performances using several statistical hypothesis tests.
- Proposed a **novel approach for extracting decision rules from SVM and Random Forest** models.
- Obtained predictive regions of large text data using CNN and developed the safety measures using Association Rule Mining.

Health Rate Prediction System

- Developed an android application that takes static and dynamic inputs from the worker and predicts their health rate.
- Collected the dynamic data from workers (heart rate) with the help of ICT based data capture system into the application.
- **Deployed a BPNN model in the android app** to predict the health rate of the worker in real-time.

Skills and Coursework

Skills	Python, R, SQL, Flask, Streamlit, C, Tableau, NoSQL, Java, Docker, Heroku, Azure ML, AWS, GCP, D3.js, PySpark, HTML, CSS, JavaScript, OpenRefine, SPSS, RapidMiner
Coursework	Deep Learning, Natural Language Processing, Machine Learning, Regression Analysis, Data and Visual Analytics, Graphical Models in Machine Learning, Time Series Analysis, Data Analytics in Business, Introduction to Analytics Modeling, Computing for Data Analysis, Business Fundamentals for Analytics, Data Structures and Algorithms

Projects

STOCK PRICE FORECASTING APPLICATION

- Predicted the buy or short signals by combining sentiment and technical analysis on real-time price movements and tweets.
- Developed an innovative & multi-faceted web application using flask and streamlit in Python.
- Deployed the application on the Heroku platform to monitor real-time forecasts.

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 Character level Convolutional Neural Network model on IMDB reviews with 87.2% accuracy.
- Developed a content-based recommendation engine on Netflix movies and TV shows data.

Publications (peer-reviewed)

JOURNALS (174 CITATIONS)

- 2022 Classification and pattern extraction of incidents: a deep learning-based approach
- 2019 An optimization-based decision tree approach for predicting slip-trip-fall accidents at work
- 2019 Application of optimized machine learning techniques for prediction of occupational accidents

CONFERENCE PAPERS (68 CITATIONS)

- 2020 Text mining-based association rule mining for incident analysis: a case study of a steel plant in india
- 2016 Study of optimized SVM for incident prediction of a steel plant in India
- 2016 Text mining based safety risk assessment and prediction of occupational accidents in a steel plant