PRATEEK GUPTA

Bangalore | +91 9916491998 | prateekg.nitjsr@gmail.com

LinkedIn| *Github*

PROFILE

Full Stack Data Scientist with close to 8 years of experience in building predictive models using machine learning & statistical concepts. Past expertise working for two Fortune 50 clients and India's largest online fashion marketplace, Myntra

TECHNICAL SKILLS

Python (pandas, NumPy, scikit-learn, Flask), SQL, Machine Learning, TensorFlow, PySpark, Cassandra, Airflow, Kafka, Docker, C++

WORK EXPERIENCE

Pharm Easy - Senior Data Scientist   **Bangalore | Jan 2019 – Present**

- Manage demand forecasting activities by leveraging CNN based Wavenet architecture with model forecast accuracy of SMAPE 0.33
- Handle NLP model for categorizing order and cancellation feedbacks using bidirectional LSTM model with 85% AUC ROC
- Improve catalog data for upstream quality checks using medicine embedding layers made with gensim
- Ensure that the data scientist department adopts the best in class processes and techniques for statistical analysis and modeling
- Train and mentor Junior Data Analysts while simultaneously guarantee statistical integrity and accuracy within the department

Myntra - Data Analyst Team Lead   **Bangalore | Jan 2016 – Dec 2018**

- Managed CEO Escalations by spearheading a team of 6 people; identified and resolved root causes for social media issues
- Created a call volume prediction model for customer support using XGBoost with 92% accuracy; used PuLP for manpower planning and optimization by formulating a linear programming model for minimizing the number of people required in shifts
- Improved payment success rate by 4% using historical transactional data of payment gateway
- Discovered leakages in return process and collaborated with multiple stakeholders to reduce customer returns by ~0.20%
- Improved network speed of supply chain by introducing dynamic thresholds for QC checks for inwarded vendor lots

Mu Sigma - Senior Business Analyst   **Bangalore | Jun 2012 – Dec 2015**

- Managed entire gamut of activities pertaining to Stakeholder Management, Sales Force Effectiveness, and Cannibalization Impact
- Directed a team of 5 people as well as maintained seamless coordination with clients and onsite to recognize business requirements and project priorities
- Established and organized QlikView dashboard for determining Sales Force Effectiveness of a Fortune 50 Pharmaceutical Company
- Discovered inefficiencies and material overages in the manufacturing process of an FDA phase 3 drug through analyzing gaps

PROJECT EXPERIENCE

- Built a NLP and machine learning pipeline to classify disaster response messages; processed text with the NLTK library using tokenization, stop-word removal and lemmatization; used TF-IDF to vectorize text and F1 score, precision and recall for evaluation
- Worked on music streaming company's data infrastructure by creating and automating a set of data pipelines with Airflow; stream public transit status using the Kafka ecosystem to build a stream processing application that shows the status of trains in real-time
- Built and trained a convolutional neural network for end-to-end driving in a simulator, using TensorFlow and Keras; used optimization techniques such as regularization and dropout to generalize the network for driving on multiple tracks
- Leveraged unsupervised learning techniques like principal component analysis and k-means clustering to compare demographics data for a German mail-order company to demographics of the German population at large
- Detected highway lane lines on a video stream; used OpenCV techniques to identify lines, including Hough Transforms and Canny edge detection; implemented a 2-D particle filter in C++ capable of localizing a vehicle within desired accuracy and time

EDUCATION

Bachelor of Technology (Hons.) Electrical Engineering | NIT Jamshedpur, 2012
Masters In Computer Science OMSCS (Online) | Georgia Tech, Fall 2020 (Incoming Graduate)
Analysis of Algorithms CS 325 (96.86%) | Oregon eCampus Post-Baccalaureate, Winter 2020
Operating System 1 CS 344 | Oregon eCampus Post-Baccalaureate, Spring 2020
Machine Learning Advanced Nanodegree | Udacity, Jun 2018
Deep Learning Specialization | Coursera, Dec 2018
Self-Driving Car Engineer | Udacity, Nov 2019
Full Stack Deep Learning Bootcamp | FSDL, 2019

MENTORSHIP

Data Scientist Nanodegree | Udacity, **Data Engineering Nanodegree |** Udacity, **Data Streaming Nanodegree |** Udacity, **AI Product Manager Nanodegree |** Udacity, **Teaching Assistant |** Springboard, **Capstone Project Mentor |** upGrad