ARKA CHOWDHURY

• (716)-986-2817 • arka.chowdhury31@gmail.com • www.linkedin.com/in/arka-chowdhury • 58-28 43rd Ave, Woodside, NY, 11377.

PROFESSIONAL EXPERIENCE

- **RSG Media Systems, New York City, NY** **Lead Data Scientist** **June 2020 – Current**
  - Designed a Machine Learning Model Pipeline for efficient and distributed Big Data ingestion and feature creation in Amazon S3 and Databricks (Apache Spark) which is used for ensembling models and predict future audience.
  - Retrieved data from Data Lakes in AWS S3 storage; joined, validated, cleaned and structured the data as an input to the ML model. The insights from the data equipped the analysts with the trends and features of audience flow.
  - Developed the Pipeline of an ensembling model that uses multiple Machine Learning algorithms for prediction modelling and automatically selects the best learning algorithm to predict the future. This model ensures the selection of best learning algorithm which has improved the prediction accuracy by 7% from the previous XG Boost model.
  - Implemented a cluster management system to optimize computation in spark by managing the cluster type, size and workers. This system reduces the execution time by 30% and saves the cost of computation.

- **RSG Media Systems, New York City, NY** **Data Scientist** **June 2018 – May 2020**
  Technology – Python, CPLEX, FICO Xpress, Apache Spark, PySpark, AWS S3, MySQL, Oracle, R
  - Developed an XGBoost model on Apache Spark using PySpark to predict the audience viewing a network at any time. The accuracy of the predictions improved the program and break schedule with better delivery of promos and advertisements, saving 12% of the advertisement inventory.
  - Created and developed an optimization model using FICO Xpress solver to improve the existing break logs for the networks by shuffling the breaks. The objective was to deliver impressions for each break as per the advertisement deals (CPM) preventing any under or over delivery saving 1.2 million dollars equivalent of break inventory each year.
  - Developed a Lasso and Elastic-Net Regularized Generalized Linear Models using GLMNET in R to forecast the audience viewing a network based on the demography, location and frequency of viewing. This model was used to derive the relationship between reach and impressions considering the factors affecting the viewers.
  - Developed a natural language processing model to extract and structure the media rights terms from the contracts. It eliminated the manual entry of legal terms ingested by the analysts, making the process instant and error proof.

- **RSG Media Systems, New York City, NY** **Data Scientist Intern** **May 2017 – April 2018**
  - Designed and developed a scheduling model in Python to get the optimal schedule of promos using CPLEX solver. It equipped campaign managers to schedule promo plans on multiple platforms (linear, digital) till the premiere date.
- **National Institute of Technology, Rourkela, India** **Research Intern** **May 2014 – July 2014**
  - Designed an algorithm to find the optimal job schedule using scatter search (heuristic for integer programming).
  - Coded the algorithm in MATLAB and the optimal schedule was proven 6% more efficient than existing algorithms.

PROJECTS

- **Enhancement of the University at Buffalo Bus (STAMPEDE) Schedule** (2018)
  Cleaned and transformed the bus GPS data to execute exploratory analysis in Tableau and list the issues in the current schedule. Collaborated with my advisor to create a new schedule which decreased the average waiting time by 12 mins, lowered the peak hour load by 200 people and enhanced the routes reducing average travel duration by 8 minutes.
- **Travelling Salesman routing problem for UPS using nearest neighbor, MILP & Simulated Annealing** (2017)
  Plotted the depots and customers in a map using Folium and MapQuest in Python. Coded the algorithms and used Gurobi as a solver for MILP. Exhibited the optimum routes and costs for the algorithms in the map.

TECHNICAL SKILLS

- **Programing Languages:** Python, R, Scala, PySpark, Mosel, Java
- **Database Tools & Languages:** MySQL, PL/SQL, Oracle, DB2, AWS S3
- **Tools and IDE:** Jupyter, Apache Spark, Minitab 17, Tableau, Power BI, MATLAB
- **Mathematical Optimization Solvers:** CPLEX, Gurobi, FICO Xpress

EDUCATION

- **University at Buffalo, The State University of New York** **Fall 2016 - May 2018**
  Master of Science in Industrial Engineering. Specialization- Operations Research
- **Veer Surendra Sai University of Technology, Burla, India** **Fall 2011 - May 2015**
  Bachelor of Technology in Production Engineering.