Sameer Shah
Data Scientist

sameerbipinshah@gmail.com
(+91) 9448173027
www.linkedin.com/in/sameershah141/

An experienced data science professional adept at building end-to-end solutions to complex business problems. Has developed expertise in cross-domain, collaborative solution building due to a strong academic background coupled with continuous learning approach. Also, has a flair for leading and managing projects with a product mindset, proactive communication and interpersonal skills while maintaining the focus on the project at hand.

Proficiency

Models & Techniques

Expert

Regression (Ridge, Lasso),

Classification (Logistic, Decision Trees, Random Forest, XGboost, LightGBM, Catboost),

Clustering (K-NN),

Optimization (Linear, Integer, Genetic Algorithm)

Intermediate

Time series (ARIMA, BSTS), Natural Language Processing, Neural Nets (Deep NN, CNN)

Technologies & Tools

Expert

Python and Jupyter,

R and R Studio,
SQL, Hive, PySpark,

Alteryx, Element,

Microsoft Office, Jira, Asana

Intermediate

Zeppelin, Hue,

Tableau, PowerBI,

Matlab, Git

Experience

❖ **Media.Net (Directi), Mumbai**
o Data Scientist   - June 2019 – Today (1.5 years)

Responsibilities:

Part of the Max-Auto-Optimization team, building accurate, scalable and robust prediction models on large data for real-time bidding, recommendation and optimization of search and display ads.

Key Projects:

❖ **Click Through Rate Prediction** – A LightGBM model to accurately predict the click-through rates of display ads. The model is used in the system that serves more than 50 million ads a day. The improvement in accuracy of the new model has estimated annual impact of $5 Million.
❖ **Conversion and Revenue Prediction** – Further steps of the CTR model, a model to predict conversion probability and revenue in $ for clicked ads. Developed and implemented models on very skewed and imbalanced data, also implemented various methods of URL encoding to further improve the accuracy.
❖ **Customized Exploration System –** Improved the parameters of the custom Exploration-Exploitation system to suit the imbalanced CTR data.

Highlights:

o Successful implementation of improved models replacing multiple legacy models
o Conducted Data Science & Machine Learning training for fellow colleagues, and Probability & Statistics training for newly joined associates.

❖ **Walmart Labs, Bengaluru**
o Data Scientist (IN-3)   - January 2018 to June 2019 (1.5 years)
o Associate Data Scientist (IN-2)   - August 2016 to December 2017 (1.5 years)

Responsibilities:

Part of the Global Data Analytics team, enabling business users to take data driven decisions by providing meaningful insights and foresights via robust statistical and machine leaning models.

Key Projects:

❖ **Store Remodel Project** – A Recommendation System to select most optimal 500 stores out of 2500 candidates for remodel by analyzing the financial and demographic aspects of the stores. This system is being used to plan and distribute $1.8 Billion of capital each year.

Sameer Shah
Data Scientist

sameerbipinshah@gmail.com
(+91) 9448173027
www.linkedin.com/in/sameershah141/

- ❖ **Customer Segmentation –** A scalable model to understand consumer behavior by analyzing monetary spends, transaction patterns across channels along with demographics details. This model is being used to make clusters on the whole US population i.e., over 325 million individuals to study various scenarios in real estate investments.
- ❖ **Modified Elo rating system for item ranking –** Inspired from the player rating in game of Chess, a modified Elo rating system to rank items in each category that are displayed on an online retail platform. This ranking system allows the category managers to implement price changes or experiment with item substitution strategies and observe the effects in real time.

Highlights:

- ❖ Filed Patent on Remodel Causal Inference and Recommendation System
- ❖ Represented Bengaluru Team in the Global Data Science Summit, US
- o Conducted internal trainings on R and Alteryx for newly joined associates

----

- ❖ **Edvancer, Mumbai**
- o Course Instructor   - January 2020 to March 2020 (3 Months)

Responsibilities:

Instructor for the CBAP (Certified Business Analytics Program) by the Edvancer Organization, took weekend classes for a batch to teach the data science and machine learning module using R in RStudio.

Academics

- ❖ **Indian Institute of Science (IISc), Bengaluru**
- o Master of Science in Mathematics   - August 2015 to July 2016 (1 year)
- o Bachelor of Science in Mathematics   - August 2011 to July 2015 (4 years)

Academic Courses:

Probability and Statistics, Probability models, Linear Algebra, Number theory, Real Analysis, Multi-variable Calculus, Combinatorics, Graph theory, Machine Learning, Cryptography, Game Theory, Spatial Dynamics in Biology.

Key Projects:

- ❖ **Simulations for Lottka-Voltera model –** Studied spatial movements of predator and prey with random walk models to determine the parameters of population dynamics.
- ❖ **Tic-Tac-Toe game –** Developed a browser based 9x9 version of the tic-tac-toe game by self-learning JavaScript as part of the KVPY summer project.

Highlights:

- o All India Rank 6 in GATE Mathematics   - May 2015
- o All India Rank 1 in NCAT (National Creativity Aptitude Test)   - March 2013
- o KVPY Fellowship for 5 years (Kishore Vaigyanik Protsahan Yojana)   - 2011 to 2016

Place –

Date –

Sameer Shah