SAHIL MAHESHWARI

(+91) 8746848619 sahil030288@gmail.com

linkedin.com/in/maheshwarisahil https://github.com/sahil-m

SUMMARY

- An engineer and an MBA with understanding of both business and technology
- Ten plus years of experience including seven plus years in data science
- Deep understanding of statistics and mathematics behind DS algorithms
- Experience in all six steps of the CRISM-DM process as a doer and also as a lead ensuring their right execution
- Understand DS project management challenges, and have experience in solving them

SKILLS

- Languages**:** R, Python, Java, C, SQL
- Understand and have worked on all major supervised and unsupervised ML (non-DL) algorithms
    - Supervised: regression and variants, tree based, bagging, boosting, SVM, Bayesian inference based, and KNN
    - Unsupervised: K-Means, HAC, GMM, and density based
- DL: Understand following architectures: CNN, RNN, LSTM, Autoencoders, GAN, YOLO, and Word2Vec, but have worked only on CNN, LSTM, YOLO, and Word2Vec. Have mostly worked on Keras, but have some experience in TensorFlow and PyTorch.
- Understand and have worked on following dimensionality reduction techniques: PCA, LDA, factor analysis, and t-SNE
- Understand and have worked Bayesian probabilistic models using PyMC3
- Have worked following model interpretability methods: PDP, LIME, and Shapley values. Have partial understanding.
- Probability and Statistics: Understand and have used
    - Probability distributions
    - Hypothesis testing and various standard tests
    - A/B testing
- Causality: Understand and have worked with do-calculus for identification and propensity-based methods for estimation.
- Reinforcement Learning: Understand SARSA, Q-Learning
- Visualization: Tableau and R-Shiny, apart from various plotting libraries in R and Python

BUSINESS EXPERIENCE

Sixt | Data Scientist | WFH from Gurgaon | Oct 2020-Present

- Fleet Planning via Simulation
  - **Implemented Bayesian demand forecaster** model to predict a distribution to sample from instead of a point estimate
  - Implemented a Multi-level Price Demand relationship model with following characteristics:
    - Discrete choice structure to model cross product price elasticity
    - **GMM to handle price endogeneity**
    - A way to inform past fleet constraints to the model
  - Implemented **GLM based multi-model prediction** model for incremental revenue.
  - Defined a **new metric to evaluate models with probabilistic output**
  - Implemented the entire simulation environment and a **Genetic algorithm-based optimizing agent** from scratch
- Add-ons Pricing
  - **Implemented causal Price Demand models** on past non-randomized data
  - **Designed and implemented an A/B test** to measure the performance of multiple pricing models
  - Designed a randomization experiment for prices
  - Measurement of price elasticity based on the randomized prices is in progress

ThoughtWorks | Data Scientist | Bengaluru | Sep 2018-Oct 2020

- Demand based pricing algorithm for a global retail client
  - Used SARIMAX and brute force optimization
  - Ensured statistical correctness of the SARIMAX model
  - **Designed AB test strategy** without having random test and control sets
  - **Defined new metrics to measure intangibles** like customer's price perception, risk appetite, etc. and used them to define a risk control strategy
  - Finally **made the price live, leading to increase in profits**
- Improving search for an intranet portal
  - Used word2Vec to get embeddings of queries
  - Used various clustering techniques to cluster queries
  - Used tf-idf scores along with boosting to match queries with documents
- Deduplication and record linkage for a consulting firm's client data
- **Active research** on how to use causality concepts and techniques like **do-calculus in ML algorithms to solve causal business problems** like marketing attribution, pricing, etc.

Happiest Minds | Data Scientist | Bengaluru | Oct 2015-Sep 2018

- Real time recommendations system for a retail client. Used **non-conventional similarity measures** resulting in more relevant recommendations
- Time series forecasting for multiple domains like energy, manufacturing, retail, etc. Handled **high frequency data (minute level)** which is usually not well handled by popular implementations of ARIMA
- Prediction of claim propensity and time to claim, for an insurance intermediary, to enable **risk based pricing**. Also, determined factors causing the claim
- Image classification using Inception model to **automate identification of damaged devices**
- Optimization of predictive model to **handle incremental data** by maintaining data preparation pipeline
- Dashboard development on R Shiny and Tableau for **actionable business insights**

Cognizant | Product Specialist | Bengaluru | Jun 2014-Oct 2015

- Responsible for business development and building analytics capabilities of the team
- Implemented **ML based credit scoring models** to replace rule based risk scorecards

Infosys | Systems Engineer | Pune | Dec 2009-Apr 2012

- Experience includes coding in Java and SQL and automation using QTP
- Built **employee summary dashboard**, leading to increased efficiency in resource management

EDUCATION

T. A. Pai Management Institute | 2012-14 | PGDM (Finance)

Apeejay College of Engineering | 2005-09 | B.E. (E.C.E.)

PRO BONO

DataKind | 2015

Predicted probability of a complaint getting resolved for Janaagraha - IChangeMyCity