Shashi Tripathi

Certified Data Science Professional

shashi.123.prakash@gmail.com | +91 8858241250 | Mumbai, IN | LinkedIn | GitHub

2+ years experience in Data Analysis & Certified Data Scientist with a passion to solve real-world business challenges using data analytics. Proficient in deploying complex machine learning and statistical modelling algorithms/techniques for identifying patterns and extracting valuable insights for the organizational leadership. Proficient in translating technical requirements into business specifications for streamlining existing processes and delivering user-centric solutions.

TECHNICAL SKILLS

- **Frameworks** NumPy, SciPy, CRISP-DM, tensorflow, Keras
- **ML/DL Techniques** k-NN, SVM, Random/Decision Forests, Logistic Regression, Linear Regression, Clustering
- **Tools & Languages** Python, R, Hive, SQL, PIG, Java, Matlab

KEY SKILLS

- Supervised/Unsupervised Learning
- Inferential Statistics
- Data Visualization & Sanitization
- Business Analysis & Strategy
- Data Mining & Data Wrangling
- Model Deployment
- Data Analysis
- Predictive Analytics & Modeling
- End to End Delivery

CERTIFICATIONS (Coursera, '18)

- Applied Text Mining – University of Michigan
- Applied SNW Analysis - University of Michigan
- Applied Machine Learning – University of Michigan
- Data Visualization and Representation - University of Michigan

KEY DATA SCIENCE PROJECTS

Domain: Acquisition Analytics (Insurance Marketing) | Mar '19

- *Objective*: Predict the probability of response of each prospect and target the ones most likely to respond to the telemarketing campaign and achieve 80% of total responders at the minimum possible cost
- *Tech Stack*: R, Python
- *Solution*: Designed a machine learning model via **logistic regression** to predict the probability of response of each prospect
- *Key Achievement:* Developed a model with an accuracy, Sensitivity and Specificity of **85, 83 and 86** respectively

Domain: College Data Clustering| Dec '17

- *Objective:* Human Resource Department was having issues in consolidating the student data with their college name because every student has it's own way of writing the college name. We used unsupervised learning to cluster the student data so that HR can identify the total number of available college and students of a particular college
- *Tech Stack*: Python, Google OpenRefine
- *Solution*: Designed **unsupervised learning based model** with **custom distance matrix** for data clustering**.**
- *Key Achievement:* Model designed to clean student data for better analytics and marketing purposes.

Domain: Help-desk Query Classification | May '18

- *Objective:* Human Resource Department was having plenty of mails every day regarding the Leaves, Holiday List, Tax and other available categories. Our task was to perform a multi-class classification on textual data and tag a category so that the process of L1 support can be automated.
- *Tech Stack*: Python
- *Solution*: Designed a machine learning model via **Linear SVC with OneVsRest** to predict the class of response of each prospect
- *Key Achievement:* Developed a model with an F1-macro of 90.

Domain: Social Network Analytics (Link Prediction) | Jun '17

- *Objective*: **Link prediction** is used to predict future possible links in the network (E.g., Facebook People You May Know Feature)
- *Tech Stack*: R

- *Solution*: Designed **decision trees** along with **random forest** to predict the probability scores also used GLM.
- *Key Achievement:* Presented the work in **Indian Institute of Technology (IIT-K), Kanpur**. Published in Springer and International Journal of Web Based Communities

KEY PUBLICATION

Hybrid Approach for Predicting and Recommending Links in Social Networks | Springer **Sep'18**

- We defined a hybrid feature-based approach that uses local graph feature by computing proximity between every pair of nodes. It also captures global feature by computing all length two and length three pathways between each pair of vertices of the network.

SimNER - An Accurate and Faster Algorithm for Named Entity Recognition | IEEE **Mar'19**

- We present a computationally more efficient approach to extract and classify the named entities. This approach uses rule based learning in combination with regular expression and pattern matching to extract the entities from the given text.

PROFESSIONAL EXPERIENCE

Data Scientist – Research and Development Division (NIA) | Tata Consultancy Services Thane, IN | Jan '18 – Present

Client: The New India Assurance, IN (PSU)

Data Analysis & Insight

- Conducting Data Analysis on Time Series Data.
- Analyzing Campaign data and present deriving factor which enhanced ROI of credit union by **20%**
- Independently conducting **data analysis** and initiated processes involving **data mining** & **data dissemination**

Software Delivery & End-User Satisfaction

- Delivered a model for fulfilling specifications & ensuring a high degree of predictability in results
- Guaranteeing **technical compatibility** & enhancing user satisfaction by closely **liaising** with **technical specialists** & end users

Documentation

- Ensuring effective documentation of systems w.r.t their purpose and function
- Performing **post-launch technical audits** as part of initiating and implementing a proactive QA process

Mentoring & Stakeholder Management

- Mentoring a team of 10 System Engineers & pacing them up for Modern Analytics

Award & Recognition

- **Received Outlier Award**

Assistant System Engineer | Tata Consultancy Services Gandhinagar, IN | Sep '17 – Dec '17

Client: Human Resource - TCS, India

ML Modelling and Analysis

- Created Multiple dashboards to monitor the traffic, Health and real time analysis of servers using **Prometheus** and **Grafana**
- Coordinated with clients and initiated measures to understand their requirements for **delivering effective solutions**
- Executed Quality Analysis for **CodeVita** applications & **collated querying results** from DB to conduct further analysis.
- Created a model to detect the software similarity and identify the plagiarized source code.

Award & Recognition

- Received **LIREL Award**

  For possessing TCS Values (Leading change, Integrity, Respect for the individual, Excellence, Learning and Sharing) and making a difference in TCS Organization among 250 associates.
- Received **TCS Innovation Pride Award**

  Provided Solution for Temporal Ordering of Clinical Data, award is given to innovative minds in TCS

EDUCATION

Specialization in Applied Data Science using Python | University of Michigan (Coursera) **Ann Arbor, MI, US | Jun '18 – Jun '19**

Master of Computer Application | University of Allahabad **Uttar Pradesh IN | Jul '14 – Jun '17**