Shashi Tripathi

Data Scientist

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

3+ years' experience in Machine Learning & 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, Pandas, CRISP-DM, TensorFlow, Keras, PyTorch, H2O, FastAI
- **ML/DL Techniques** Linear Modeling, Ensemble Modeling, CNN, RNN, LSTM, Transformers, Transfer Learning
- **Tools & Languages** Python, R , Java, Matlab, C, C++
- **Cloud Platforms & REST** AWS, Azure, Tata Cloud, Heroku, Flask, FastAPI,
- **Edge Devices Frameworks** OpenVINO, TensorFlow Lite

KEY SKILLS

- Supervised/Unsupervised Learning
- Inferential Statistics
- Information Retrieval
- Natural Language Processing
- Edge Computing
- Model Deployment
- Computer Vision
- 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
- Applied Data Science Specialization - University of Michigan

Contribution and Ownership (Open Source)

- NetworkX
- TextVectorizer: Representation Learning for Text
- Tensechanger: Converts tense of a sentence
- Data Explorer: Data Visualization App

KEY DATA SCIENCE PROJECTS

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 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, Python
- *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

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: 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

KEY PUBLICATION

Hybrid feature-based approach for recommending friends in social networking systems | IJWBC Geneva, CH| Feb'20

- We designed two time efficient algorithms for finding all paths of length-2 and length-3 between every pair of vertices in a network which are further used in computation of final similarity scores in the proposed method. Further, we define a hybrid feature-based node similarity measure for link prediction that captures both local and global graph features. The designed similarity measure provides friend recommendations by traversing only paths of limited length, which causes more faster and accurate friend recommendations. Experimental results show adequate level of accuracy in friend recommendations within considerable computing time.

SimNER - An Accurate and Faster Algorithm for Named Entity Recognition | IEEE Allahabad, IN| 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.

Hybrid Approach for Predicting and Recommending Links in Social Networks | Springer Kanpur, IN|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.

Applying Model View View-Model and Layered Architecture for Mobile Applications | JIAPS Allahabad,IN|Jul'16

- The paper applies the Model View View-Model for mobile applications. The exponentially increasing user of mobile devices is leading to a next generation of applications that exploit user contextual information to provide a richer experience. The activities to perform during the development of application may vary on their context. The paper lays emphasis on the layered architecture of mobile applications. The methodological process followed during the ontology development as well as the ontology platform obtained from this process. Besides, we provide an example of how to extend the ontology for a particular use case in a concrete domain

PROFESSIONAL EXPERIENCE

Machine Learning Engineer – Research and Innovation (A&I) | Tata Consultancy Services Pune, IN | Jan '20 – Present

Working on ADD (TCS Product).

Tata Consultancy Services' (TCS') Advanced Drug Development (ADD) is a suite of platforms for the entire clinical R&D value chain. Our platforms harness new digital technologies such as automation, AI, and IoT–to bring about disruption in clinical research and integrate key features of our intellectual property (IP) assets with commercial-off-the-shelf offerings.

- Textual Data Modeling
- Pattern Recognition and Rule Generation
- Human Reading Mimicking
- Template-less Information Retrieval
- Knowledge graph and Ontology Building from Text

Data Scientist – Research and Development | Tata Consultancy Services Thane, IN | Jan '18 – Dec' 19

Client: Confidential

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 | Dec'18 – Jun '19**
Master of Computer Application | University of Allahabad **Uttar Pradesh IN | Jul '14 – Jun '17**