

Tarun Kumar, IIT Kanpur
LinkedIn: <https://www.linkedin.com/in/kumartarun123/>

Email Id: tarunjnv1995@gmail.com
Ph No. +918004240842

TECHNICAL SKILLS

Data Science: Python, Machine learning Algorithm, Deep learning (Beginner), NLP

Tools: Jupyter notebook, Spyder, Postman, VS-Code, Big-query, APIs

Cloud: Google Cloud Certified Associate Cloud Engineer

PROFESSIONAL EXPERIENCE

Data Scientist | xLABS (AI/ML & Cloud Research team), Virtusa (2018-2020)

Cohort-Sense Analyzer (GCP-Solution) (Jan'20-Apr'20)

- Integrate and analyze subject and data together using innovative visualization and analysis tools. It's an AI-driven simulated data set generated based on North-America statistics region through SYNTHETA (Synthetic data generation tool). Deployed it on **GCP** on Kubernetes engine

Audio Emotional classifier using Deep learning | xLABS (Dec'19-Jan'20)

- Developed a model using Machine learning to classify emotions of audio data. The work-flow were 4 stages: Data Exploration using graphical representation, extracting audio features using MFCC, apply ML algorithm and classify the test audio using our finalized model

Entity Extraction | Client: New York University (Oct'19-Nov'19)

- Automated a complete work flow for named entity extraction from parsing pdfs to automatically getting desired entities. The work flow were three stages; the text extraction from pdf and images using **tesseract** tools, classification of the text-document using **AI/ML** technique, and finally desired named entities extraction using **Stanford NER** and show it on Postman

Insurance Lapse Prediction | Client: IBM (July'19-Sep'19)

- Analyzed the insurance data using EDA and understood the important features using hypothesis testing and correlation matrix and followed by over-sampling techniques (**SMOTE**)
- Advance Machine learning to train a model to predict whether a policy is likely to lapse or not
- Evaluate & contrast the performance of using evaluation parameters such as AUC & ROC curve

RegTech: Automate Regulatory form using knowledge graph | Client: CITI Bank (Mar'19-June'19)

- Discover the data-relationship of the regulatory forms metadata using knowledge graph & explore across a range of business requirement using Neo4j, **NLP** and graph ontology
- Achieved 50% cycle time reduction for new country/region form onboarding by **semantic similarity**

RegInsight: Data quality improvement by AI/ML | Client: CITI Bank (Nov'18-Mar'19)

- Provided a solution to improve the data quality of RegInsight using **EDA, feature engineering & data Analysis, Association rule mining (Apriori)/ Combination Analysis (MBA-market basket analysis)** & Using **k-means clustering** algorithm for anomaly detection
- Collaborate with team to deploy models as APIs consumable by UI to drive business values

Academic Qualification & Personal Information

• **Indian Institute of Technology Kanpur**, Kanpur (U.P), India (Jul'14-Jun'18)
B.S CHM & MINOR-(MBA/IME)

• **Jawahar Navodaya Vidyalaya – Mau** (U.P), India (Jul'06-Apr'13)
6th- 12th

Relevant Courses

- **Coursera:** Introduction to data science in Python, Machine Learning & Deep Learning Specialization by Andrew Ng (*Ongoing)
- **Pluralsight Certifications:** Certifications for 25 Statistics, Machine learning & deep learning courses
- **Stanford courses:** CS231n (Convolutional neural network for visual recognition), CS224n (Natural Language Processing with Deep Learning) (*Ongoing)
- **Linux Academy:** Google Cloud Certified Associate Cloud Engineer