Kapil Sahu

Boston, MA | (+1) 857-867-2861

kapilsahu2102@gmail.com | www.linkedin.com/in/kapilsahu- | https://kapilsahukp.github.io

EDUCATION

Northeastern University, Khoury College of Computer Sciences, Boston, MA   (Expected) May 2023

Master of Science, Data Science

Indore Institute of Science and Technology, India   August 2012 – July 2016

Bachelor of Engineering, Computer Science and Engineering   **(GPA: 8.03/10)**

TECHNICAL SKILLS

Programming Languages: Python, SQL, R, JAVA, C++, C, HTML.

Frameworks and Libraries: Pandas, NumPy, Scikit-learn, NLTK, Spacy, Matplotlib, Seaborn, Git, RDBMS, MySQL, PostgreSQL, DataGrip, PyTorch, Keras/Tensorflow, Tableau, AWS, Google Cloud Platform (GCP), QGIS, MS-Excel.

Knowledge Domains: Data Structures, Algorithms, Linear Algebra, Statistics, Probability, Natural Language Processing, OOP Design, ERD, Data Modeling, Machine Learning, Deep Learning, Regression, IR, A/B Testing, ETL, CI/CD.

PROFESSIONAL EXPERIENCE

Charles River Data (Data Scientist, Intern)

Revenue and Finance Management: May 2022 – Present

- Reduced latency by **40%** by redesigning and implementing end-to-end data consumption, processing, and analysis pipeline for historical financial data (insurance payments) from multiple API endpoints.
- Enhanced work efficiency by **70%** by designing Tableau dashboard for business to check sales by location, products, and providers.

Flood Insurance Premium Prediction and Anomaly Detection: May 2022 – Present

- Improved prediction accuracy by **30%** by incorporating parameters from National Flood Insurance Prog. using **Decision Trees** for feature selection and using **PyOD** for outlier detection.
- Reduced payment risk factor by **20%** after analyzing geographical data points of property using QGIS.

Geocoding Error Detection: May 2022 – August 2022

- Engineered geographical features to improve efficiency of detecting the geocoding errors by **30%** in residential and commercial properties using **Zillow** datapoints, Open Street Maps and QGIS.

Zensar Technologies/ LorhanIT Services Pvt. Ltd. (Software Engineer, **Data Analytics)**

Commercial Aviation Crew Leave and Payroll Management System: December 2016 – October 2020

- Reduced workload by **60%** after analyzing **financial data** and developing automated payroll mechanism.
- Enhanced user experience over **30%** by introducing a feedback mechanism in the Flight Plan (iPad app).
- Lead client-vendor discussion at multiple stages of project and trained **6** resources with business knowledge.

PERSONAL AND ACADEMIC PROJECTS

Salary Predictor:

- Created an end-to-end salary predictor by training the model on scraped LinkedIn and Glassdoor data.
- Model implementation **(Multiple Linear Regression, Lasso Regression, Random Forest).**

Sentiment Analysis (Sarcasm Detection): (Accuracy = **72.15%** approx.)

- Sarcasm detection using CNN, RNNs like **LSTMs** and making use of techniques like **bag-of-words** with **TF-IDF**, **Word2Vec** and/or **GloVe** to predict whether a comment is sarcastic or not.

Spam SMS Detector/Classifier: (Accuracy = **97.21%** approx.)

- Created a classifier to detect whether an SMS is spam or legit by using data modeling techniques like **bag-of-words**, **TF-IDF** and **Naïve Bayes Classifier** to predict whether an SMS is spam or not.

Question Answer Model: (Accuracy = **70.9%** approx., F1-Score = **80.95**)

- Implemented Information Retrieval and transformers **(BERT, DistilBERT, ALBERT & Ensemble)** to create a Question-Answering Model based on SQuAD1.1 dataset.

INTERESTS

- Meditation, Running, Cooking, Volunteering for cleanliness and sustainable living awareness drives.