Kapil Sahu

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

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

Professional Experience

Data Scientist, Intern **May 2022 – Dec 2022**

***Charles River Data**, Boston, MA*

Flood Insurance Premium Prediction, Anomaly Detection & Geocoding Error Detection:

- Saved **$600K** per quarter by **improving prediction** accuracy by **30%**, using National Flood Insurance Program parameters, implementing **Decision Trees** for feature selection, and using **PyOD** for outlier detection.
- Minimized payment risk factor by **20%** after analyzing **geographical data** points of property using **QGIS**.
- **Engineered geographical features** to improve efficiency of detecting geocoding errors by **30%** in residential and commercial properties with **Zillow** datapoints, **Open Street Maps** and **QGIS**.

Revenue and Finance Management:

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

Software Engineer, Data Analytics **Dec 2016 – Oct 2020**

Zensar Technologies/ LorhanIT Services Pvt. Ltd.

Commercial Aviation Crew Leave and Payroll Management System:

- Reduced manual workload by **60%** after analyzing **financial data** and developing automated payroll mechanism.
- Enhanced user experience by **30%** on introducing a feedback mechanism in Flight Plan (iPad app).
- **Lead** project meetings with client and **cross-functional teams** at multiple stages of product development.
- Trained a team of **6** resources with **Aviation and Financial** domain business knowledge.

Education

Masters, Data Science **Sep 2021 – Aug 2023**

Northeastern University, **Khoury College,** Boston, MA **GPA:** 3.6/4.0

Bachelor of Engineering, **Computer Science** **Aug 2012 – Jul 2016**

Indore Institute of Science and Technology, India **GPA:** 8.03/10.0

Technical Skills

Programming Languages: Python, SQL, R, 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-Office.

Knowledge Domains: Data Structures, Algorithms, **Natural Language Processing**, OOP Design, Database Design, Machine Learning, Deep Learning, Regression, IR, A/B Testing, CI/CD, **Data Modeling**, **Data Manipulation**, **ETL**.

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%**)

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

FakeCheck (Image Forgery Detection): (Accuracy = **97%**)

- Developed an image classifier to detect whether an image of human face is REAL or FAKE using **CNN**, **VGG** and **DenseNet** Deep Learning Models .

Question Answer Model: (Accuracy = **71%**, F1-Score = **81%**)

- Implemented Information Retrieval concepts 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.