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:

- Achieved **$600K/quarter** savings for the client by reducing feature space of National Flood Insurance Program parameters.
- Improved accuracy by **30%** via **Decision Trees** implementation for feature selection and using **PyOD** for outlier detection.
- Reduced payment risk and boosted geocoding efficiency by **20-30%** through ranking and engineering geospatial data points for residential and commercial properties.
- Conducted **data analysis**, research, POC and designed ML pipeline along with **presenting key insights** to the stakeholders.

Revenue and Finance Management:

- **Redesigned** payments processing pipeline to achieve **40% latency reduction** by integrating multiple API endpoints.
- Optimized API hit rate to less than **1K hits/day** from 5K hits/day saving the client **$6000** of premium subscription.
- Boosted work efficiency by **70%,** saving **720 hrs.** of work by developing an executive-facing **Tableau dashboard** and identifying key sales metrics to track business progress and initiatives.

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 using SSRS
- Enhanced user experience by **30%** on introducing a feedback mechanism in Flight Plan (iPad app).
- **Led** client meetings and collaborated with **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, Hadoop, 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, **Machine Learning**, Database Design, 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.