Kapil Sahu

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

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

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

Professional Experience

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

Charles River Data, Boston, MA

Lumber Price Optimization:

- Built **custom word embeddings** to compare items from multiple stores based on **semantic similarity measures**.
- Optimized purchase cost by **40%,** implementing the Name Entity Recognition (**NER**) to identify similar items.

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

- Achieved **$600K/quarter savings** by reducing feature space of National Flood Insurance Program parameters using **QGIS**.
- 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 **features**.
- Conducted **multidimensional data analysis, 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

Commercial Aviation Crew Leave and Payroll Management System:

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

Technical Skills

Programming Languages & Databases:	Python, Java, Scala, C++, C, HTML/CSS, **SQL**, NoSQL
Frameworks and Libraries:	Tensorflow/Keras, Pandas, NumPy, Scikit-learn, NLTK, Spacy, Matplotlib, **Spark**
Proficient in Tools and Topics:	Git, **Docker**, REST API, **AWS**, Google Cloud Platform (**GCP**), **Tableau,** Algorithms, OOP Design, **Natural Language Processing**, Database Design, **Machine Learning**, CI/CD, **Deep Learning, A/B Testing**, Predictive Modelling, **ETL**, Statistical Analysis.

Personal and Academic Projects

Salary Predictor:

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

Sentiment Analysis (Sarcasm Detection):

- Implemented RNNs, including **LSTMs,** and utilized **encoding techniques** such as bag-of-words with TF-IDF, Word2Vec, and GloVe to detect sarcasm in comments with an **accuracy of 72%**.

FakeCheck (Image Forgery Detection):

- Developed and deployed an end-2-end image classifier on GCP to detect forged images and FAKE faces using **Streamlit** for UI design and **CNN**, **VGG** and **DenseNet** Deep Learning Models to classify with an **accuracy of 89%.**

Question Answer Model:

- Implemented Information Retrieval model and transformers **(BERT, DistilBERT, ALBERT Ensemble)** to create a Question-Answering Model based on SQuAD1.1 dataset, predicting correct answers with **71% accuracy** and **81% of F1 Score**.