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, Matplotlib, Seaborn, RStudio, Git, MySQL, Microsoft SQL Server, SQLite, Oracle SQL, DataGrip, PyTorch, Keras, Tableau, AWS, QGIS, MS-Excel.

Knowledge Domains: Data Structures and Algorithms, Linear Algebra and Statistics, NLP, OOP Design, Supervised/Unsupervised ML, Deep Learning, Neural Networks, Regression, Probability Theory, Information Retrieval, A/B Testing.

PROFESSIONAL EXPERIENCE

Charles River Data (Data Scientist, Intern)

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

- Improved prediction accuracy by 3**0%** 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.

Lumber Pricing Project: May 2022 – Present

- Reduced manual task of product selection by **80%** with optimized construction material cost calculator. Achieved goal of enhancing the user experience by over **50%** on introducing the shopping cart feature.

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

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

- Reduced workload by **60%** after analyzing and developing automated flight reporting mechanism.
- Enhanced user experience over **30%** by introducing a feedback mechanism in the Flight Plan (iPad app).
- Lead client-vendor discussion for requirement gathering and trained **6** resources with business and application 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.

Miniposter Representing Goodreads Books EDA:

- Created a Mini-Poster as a part of Academic project displaying interesting insights about popular books.
- Environments: **RStudio**.

INTERESTS

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