KABIR THAKUR

(571) 591-8688 • kabir.thakur416@gmail.com • kabirthakur.github.io/portfolio/ • LinkedIn

EDUCATION

Syracuse University, School of Information Studies, Syracuse, NY   May 2024
M.S. Applied Data Science
Relevant Coursework: NLP | Machine Learning | Big Data Analytics | Artificial Intelligence | Text Mining with LLMs | AWS

Central University of Punjab, Department of Computational Sciences, Punjab, India   May 2022
M.S. Physics (Computational Physics)
Relevant Coursework: | Python programming | FORTRAN | Mathematics for Computational Sciences | Scientific Programming

Shiv Nadar University, Department of Physics, NCR, India   May 2018
B.S. Physics (Research)
Relevant Coursework: Linear Algebra | Calculus II | Data Management and Analytics | Python for Physics | Intro to Programming (C)

Skills

Programming Languages: Python, R, SQL (MS SQL, MySQL), Bash scripting, C++, C, MATLAB, Octave, FORTRAN, REST API
Machine learning: Fine Tuning LLMs, Deep Learning (CNN, RNN, LSTM), Regression, Random Forests, LGBM, XGBoost, Time series forecasting, SVM, Decision Trees, kNN, Transfer Learning, Monte Carlo Simulations, Huggingface, MDP, MCTS
Data Management & Visualization: AWS, GIT, Tableau, PowerBI, Snowflake, Docker, Excel
Libraries: TensorFlow, PyTorch, Pandas, scikit-learn, PySpark, NumPy, NLTK, Spacy, ggplot2, dyplr, caret, matplotlib, seaborn

WORK EXPERIENCE

Tutor for Student Athletes – Stevenson Educational Center, Syracuse University   Aug 2023 – Present
- Tutoring undergraduate student athletes in courses on Data Analytics in R, probability, statistics, and calculus.
- Provided academic support in compliance with NCAA guidelines.

PROJECTS

Skillspotter: Named Entity Recognition on Job Descriptions   Sep 2023 - Present
Project Lead (Python, PyTorch)
- Created a dataset of 100000+ rows by web scrapping job portals. Cleaned and tokenized job descriptions for BERT model.
- Created a list of 8000+ soft and tech skills. IOB tagged skills using pattern matching and regular expressions.
- Transformed data for training BERT model to perform Named Entity Recognition (NER) with 60:20:10 for training, validation, and test.
- Trained a distilbert-base-cased model from HuggingFace to **identify skills from job description**s achieving 93% 3-fold-CV.

Yelp Recommendation Engine   Mar 2023 – May 2023
Project Lead (Python, PySpark)
- Led a team of 4 to develop a recommendation engine using Yelp dataset and PySpark.
- Performed data cleaning and transformations followed by feature engineering to implement **K-means** and **ALS** algorithm.
- Developed scalable framework to recommend 2 similar restaurants for each restaurant and 2 similar users to each user.
- Increased the number of relevant recommendations by 60% through the integration of a hybrid K-means and ALS model.

Dynamic Human AI Collaboration (ICLR Publication)   Feb 2023 – Mar 2023
Decision Science Team at JP Morgan Chase, London (Python)
- Collaborated with a 3-member team to optimize algorithmic decisions through expert integration using a Bayesian framework.
- Conducted 5 experiments with the SA Heart dataset, optimizing expert-algorithm communication in various scenarios.
- Set a baseline using independent RF models, achieving a notable AUC ROC boost to 78.91% with our Deferral system.
- Attained a top F1 score of 64.86% using "Information Sharing with dynamic priors", highlighting effective expert integration.

LEADERSHIP EXPERIENCE

QuantumCuse, Quantum Computing Club, Syracuse University
Director of Education
- Spearheaded the creation of educational resources and reusable modules for quantum computing beginners.
- Developed versatile modules specifically tailored for integration in upcoming hackathons.