KABIR THAKUR

(571) 591-8688 • kathakur@syr.edu • kabirthakur.github.io/portfolio/ • LinkedIn

EDUCATION

Syracuse University, School of Information Studies, Syracuse, NY   May 2024
M.S. Applied Data Science (Specialization: NLP)
Relevant Coursework: NLP | Machine Learning | Big Data Analytics | Artificial Intelligence | Text Mining with LLM | 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: Deep Learning (CNN, RNN, LSTM), Torch, Reinforcement Learning, Regression, Random Forests, LGBM, XGBoost, Time series forecasting, SVM, Decision Trees, kNN, Transfer Learning, Monte Carlo Tree Search, Huggingface, MDP
Data Management & Visualization Services: AWS (S3, EC2, SageMaker, Lambda, Glue),GIT, Looker Studio, Tableau, PowerBI
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
- Tutored 12 undergraduate student athletes in courses on **Data Analytics in R, probability, statistics, and calculus**.
- Facilitated an average grade improvement of 25% among tutored students by developing tailored learning strategies.

PROJECTS

Skillspotter: Named Entity Recognition on Job Descriptions   Sep 2023 – Dec 2023
Python, PyTorch, NLU, NLI
- Created a dataset of 100K+ rows by web scrapping job portals. Cleaned and tokenized job descriptions for **BERT** model.
- Built a taxonomy of 8000+ soft and tech skills. **IOB tagged** skills using pattern matching and regular expressions.
- Trained a distilbert-base-cased model from HuggingFace to identify skills from job descriptions achieving 98% accuracy.
- Cumulated 34 sets of required skills for different tech roles and built a recommendation engine based on similarity score.

Yelp Recommendation Engine   Mar 2023 – May 2023
Python, PySpark
- Led a team of 4 to develop a recommendation engine using Yelp dataset and Spark.
- Cleaned and transformed 1M+ rows of data 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.

HealthCost Insight: Reducing Healthcare Cost   Mar 2023 – May 2023
RStudio, dyplr
- Led a 4-member team to pinpoint primary expense drivers, resulting in a 20% cost reduction for an HMO.
- Managed extensive data cleaning, segmenting the dataset at the 75% cost quantile for precise binary classification.
- Implemented 3 ML models (**Linear Regression, Tree Bag, SVM**), boosting predictive accuracy by 15% for healthcare costs.
- Designed an interactive Shiny App Dashboard for 4 types of visualizations-Histograms, Scatterplot, Boxplot, Map Plots

LEADERSHIP EXPERIENCE

IIT2024 Global Conference, Washington DC   Jan 12-14, 2024
Lead Volunteer Team
- Helped organize a team of 50 volunteers to manage over 1500 attendees for a 3-day conference in Washington D.C.
- Managed LinkedIn and Instagram for the conference, growing social media outreach by over 50%.

QuantumCuse, Quantum Computing Club, Syracuse University   Jan 2023 – Dec 2023
Director of Education
- Spearheaded the creation of 5 educational resources and 5 reusable modules for quantum computing beginners.
- Organized workshops for 20+ members, resulting in a 25% increase in club membership due to enhanced engagement.