KABIR THAKUR

(571) 591-8688 • kathakur@syr.edu • linkedin.com/in/kabir-thakur • github.com/kabirthakur

EDUCATION

Syracuse University, School of Information Studies, Syracuse, NY — May 2024
M.S. Applied Data Science
Relevant Coursework: NLP | Machine Learning | Big Data Analytics | Quantitative Analysis for Data Science | Text Mining | DBMS

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 | Overview of Data Management and Analytics | Algebra I

PROJECTS

ASL Alphabet Recognition — Apr 2023 - Present
- Developed a deep learning model (CNN) for ASL alphabet recognition using TensorFlow.
- Created an efficient Convolutional Neural Network (CNN) architecture for the image recognition task.
- Performed transfer learning using ResNet and EfficientNetB3 to compare and improve performance.
- Assessed model performance using f1 scores and performed rigorous hyperparameter tuning to optimize the model.

Yelp Recommendation System — Mar 2023 – May 2023
- Led a team of 4 to develop a recommendation engine using Yelp dataset and PySpark.
- Performed extensive data preprocessing and feature engineering, including cleansing, transformations, aggregations, and feature extraction from large-scale data.
- Employed NLP and Named Entity Recognition (NER) techniques to improve feature set.
- Implemented K-means clustering and collaborative filtering through ALS model.

Dynamic Human AI Collaboration — Feb 2023 – Mar 2023
- Collaborated with the Decision Science Team at JP Morgan Chase, London to work on creating a framework to supplement algorithmic decision making with expert opinions.
- Conducted experiments with the SA Heart dataset, optimizing expert-algorithm communication in various scenarios.
- Demonstrated that a Bayesian framework with priors for human(expert)-in-the-loop pipelines shows superior performance.
- Paper accepted with invite to present at ICLR 23 – Tiny Papers.

Spam Detector — Dec 2022 – Jan 2023
- Utilized the publicly available Enron email corpus to develop and evaluate multiple spam classification models using NLP.
- Employed advanced feature engineering techniques, including bag of words, POS tagging, bigrams, trigrams, and subjectivity analysis to enhance the accuracy of classifiers.
- Leveraged Python's regex library to design custom patterns to effectively identify spam-related keywords and phrases.
- Performed comprehensive model evaluation, including cross-validation and key metrics, to ensure classifier robustness

TECHNICAL SKILLS/CERTIFICATIONS

Programming Languages: R, Python, Bash scripting, FORTRAN, MATLAB, Octave, C, C++
Others: MS SQL, MySQL, Latex, ML Models (Logistic Regression, Decision Trees, SVM, kNN), Deep Learning(CNN, RNN), Google Analytics, Tableau, PowerBI, AWS, GCP, Excel (lookup, correlation, linear and nonlinear regression, forecasting), Statistical testing
Libraries: PySpark, NLTK, Spacy NumPy, Pandas, matplotlib, seaborn, scikit-learn, TensorFlow, PyTorch, ggplot2, dyplr, caret

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.

Student Career Services, Central University of Punjab
- Served as the proactive departmental student representative for postgraduate students, liaising with career services.
- Collaborated with the university to host career workshops for students.