Yelena Y. Yu

(818) 960-5029 yu.yue16@northeastern.edu yelena.info linkedin.com/in/yelenayu

Technical Skills

Programming Languages: Python, Java, C/C++, JavaScript, HTML/CSS, SQL, R, OOD/OOP
Web Frameworks: React.js, Flask, Django, D3.js, Spring Boot, Node.js, JQuery
Machine Learning Frameworks: scikit-learn, SKtime, AutoGluon, PyTorch, Keras, XGBoost, Spark, Numpy, Pandas
Databases: Amazon RDS, Amazon DynamoDB, Amazon RedShift, MongoDB, Redis, MySQL, PostgreSQL
Tools and Technologies: Git, Linux, Docker, Virtual Machines, RPC, REST, Amazon SageMaker

Experience

Arcadia.io May 2024 – Present
Data Scientist Intern - AI/ML for Health Forecasting

- Automated data processing workflows in AWS RedShift for data cleaning and pre-processing, managed processed data in Amazon S3, reducing manual data handling time from 1 hour to 5 minutes with reusable SQL queries.
- Employed PySpark to calculate statistical measures; used Matplotlib for exploratory data analysis and data visualization.
- Developed machine learning algorithms with scikit-learn in AWS SageMaker, using K-mean and KNN for clustering, multivariate time series forecasting, XGBoost, leveraging feature engineering to boost model performance.
- Evaluated predict accuracy using Area under the curve (AUC), improved respiratory disease prediction accuracy by 17%.
- Streamlined project management by integrating Jira and Confluence, collaborated with cross-functional teams to align goals and vision, improved team collaboration, and reduced time by 15%.

Research

Arcadia.io & Northeastern University Sep 2024 - Present
Deep Learning for Disease Prediction - Python, PyTorch, RNN-LSTM/GRU, Transformer

- Implementing an RNN-GRU model on our unique dataset with vast patient information to assess improvement it provides compared to existing researches using smaller, medium-sized datasets.
- Evaluating performances of RNN-GRU and Transformer models using AUC, focusing on the degree of improvement achieved by using Transformer for cardiovascular disease prediction, particularly in heart failure risk.

Open-source

SKTIME Python ML and AI Framework for Time Series
Summer Mentorship Program || Open-Source Contributor Feb 2024 – Present

- Developed an outlier detection class for time series data by implementing a sliding window approach, utilizing modified Z-score, clustering techniques, and k-nearest neighbor(KNN) algorithms.
- Maintained estimator wrappers to ensure compatibility with the sktime interface.
- Participated in the Summer Mentorship Program, working with mentors to design and develop scalable software architecture for benchmarking framework, including statistical evaluation, visualization, and reporting functionality.
- Created pull requests to enhance the library performance and address open issues.

Selected Projects

What to Watch - ***R, Shiny, Machine Learning, Collaborative Filtering*** Mar - Jun 2024

- Developed both user-based and item-based collaborative filtering models using R, employed dataset from MovieLens, and implemented dual recommendation functions: user-specified genres and user ratings to enhance personalization.
- Designed an intuitive web interface using Shiny framework for streamlined movie selection and recommendations.
- Solved the cold-start problem by requiring new users to rate a minimum of 10 movies. Otherwise, initial recommendations are based on the platform's top 10 most popular movies.
- Reduced the Root Mean Square Error (RMSE) from 1.41 to 0.97, utilizing an 80/20 train-test split.

Cloud9 Café - ***Amazon Cloud Service (AWS)*** Jan - Apr 2024

- Created and hosted a dynamic web application for a coffee shop with ordering and billing systems using Amazon EC2, utilizing AWS S3 for storing static assets and backup data.
- Implemented auto-scaling with load balancer for high availability and performance across different availability zones.
- Configured IAM roles for different employees to manage access and set up a relational database using Amazon RDS.
- Utilized VPC, NAT, Internet Gateway to enhance security and isolation, automated backup for disaster recovery.

yelena.info - ***React.js, JavaScript, Bootstrap, MaterialUI, CSS, HTML*** Jun - Jul 2023

- Designed and developed a personal portfolio app using JavaScript and React web framework.
- Used Bootstrap and MaterialUI to create dynamic elements and smooth transitions.

Police Fairness Modeling and Analysis - ***Police Fairness Analysis | R, Statistical Modeling, Data Science*** Mar - Apr 2023

- Developed a hierarchical Bayesian regression model to investigate discrimination against Asians during traffic stops.
- Implemented R/JAGS to clean and preprocess data, train, run Markov Chain Monte-Carlo simulations.
- Identified clear evidence of discrimination against Asian drivers by comparing regression coefficients of race indicators, with a statistical significance score of 1.31 for Asians compared to 0.29 to 0.97 for other races.

Education

Northeastern University Jan 2023 – Present
Master of Science in Computer Science (GPA 4.0)

Courses/Certificates:
Python, Discrete Structures, Object-Oriented Design (Java), Data Structures and Algorithms (Python & C/C++), Cloud Computing (AWS), Web Development, Machine Learning with Python, Deep Learning for Healthcare, Data Mining