Yelena Y.

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

Education

Northeastern University Jan 2023 – Present
Master of Science in Computer Science (GPA 4.0)
Courses/Certificates:
Discrete Structures, Object-Oriented Design, Data Structures and Algorithms, Cloud Computing (AWS), Web Development, Machine Learning, Deep Learning for Healthcare, Data Mining, Natural Language Processing

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, reduced data handling time from 1 hour to 5 minutes with reusable SQL queries.
- Used PySpark for statistical measure calculation and Seaborn for exploratory data analysis and visualization.
- Developed machine learning models in AWS SageMaker using scikit-learn, applying K-means and KNN for clustering, multivariate time series forecasting, and XGBoost for predictive modeling, performed feature engineering to enhance performance.
- Evaluated prediction accuracy using area under the curve (AUC), improved respiratory disease prediction accuracy by 17%.

Research

HAI Lab -Northeastern University Jan 2025 - Present
Graduate Researcher
- Conduct research and write papers on large language model (LLM) and natural language processing (NLP) research projects.
- Participated in reading groups, peer-reviewing and providing feedback on drafted papers, cross-checking data and results.

Arcadia.io & Northeastern University Sep 2024 - Dec 2024
Graduate Researcher - Deep Learning for Heart Failure Prediction
- Designed novel architectures for heart failure prediction, enabled integration of time-series vital data with demographic info.
- Implemented RNN-GRU, RNN-LSTM, Transformer in PyTorch, successfully forecast heart failure for varyious future time windows.
- Performed hyperparameter and architecture tuning using Optuna, achieving highest AUC of 0.87.
- Earned 2nd place at Poster Day among 100+ research projects.

Open-source

SKTIME Python ML and AI Framework for Time Series
Summer Mentorship Program || Open-Source Contributor Feb 2024 – Present
- Developed outlier detection for time series data using sliding window, modified Z-score, clustering, and k-nearest neighbor (KNN).
- Participated in the Summer Mentorship Program, worked with mentors to design a benchmark framework with documentation.

Selected Projects

Large Language Model (LLM) for Clinical Notes - ***Llama 3.2, Python*** Aug - Oct 2024
- Integrated LLM agent system to convert unstructured clinical notes to structured, standardized medical ICD-9 and ICD-10 codes.
- Used this data to improve performance of existing deep learning and statistical models, integrated new workflow into ML pipeline.

What to Watch - ***R, Shiny, Machine Learning, Collaborative Filtering*** Mar - Jun 2024
- Created user-based and item-based collaborative filtering model. User-specified genres and ratings to enhance personalization.
- Designed and developed an intuitive web interface using Shiny framework for a streamlined and seamless UI/UX.
- Used feature engineering techniques, solved cold-start problem by requiring new users to rate a minimum of 10 movies.

Cloud Lens - ***Java, Kubernetes, Docker, Amazon RDS, Kafka, Spark*** Jan 2024 - Present
- Developed scalable, cloud-native microservices with RESTful interfaces with Java, used AWS for robustness and scalability.
- Engineered a Reddit analytics service capable of processing 1.4TB dataset using Amazon RDS, Kafka data streams, Spark for efficient distributed operations to achieve scalability, handling up to 30,000 requests per second.
- Used Docker for service containerization and Kubernetes for orchestration, ensuring seamless deployment and management.

Cloud9 Café - ***Amazon Cloud Service (AWS)*** Jan - Apr 2024
- Created and hosted a dynamic web app for a café with ordering and billing on AWS EC2. Static assets, backups stored on AWS S3.
- 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.

Technical Skills

Programming Languages: Python, Java, C/C++, JavaScript, HTML/CSS, SQL, R
AI/ML Frameworks: scikit-learn, SKtime, AutoGluon, PyTorch, Tensorflow, Keras, XGBoost, Spark, SageMaker, Optuna, Ray
Databases: Amazon RDS, DynamoDB, RedShift, MongoDB, MySQL, PostgreSQL
Tools and Technologies: Git, Docker, Kubernetes, Hadoop