Shuowen Wei | Machine Learning Engineer
weisw9@gmail.com | (314) 215-8348 | LinkedIn: www.linkedin.com/in/shuowenwei

Summary

Solutions-driven, self-motivated machine learning engineer with 8+ years of experience in deep learning, natural language process (NLP), traditional machine learning, big data engineering, full-stack software design and development. Strong background in mathematics and computer science, extensive hands-on experience with multiple programming languages and big data platforms. Expertise in transforming business resources and requirements into manageable data formats, building end-to-end data ETL and model deployment pipelines at scale, visualizing business findings and insights, bringing excellent problem-solving skills to efficiently and creatively resolve complex issues.

→ Deep Learning, Machine Learning, NLP (Transformers, BERT, XLNet, RoBERTa etc), Graph Neural Networks.
→ Python, PyTorch, Keras/TensorFlow, Power BI, R, Matlab, SAS.
→ AWS EC2, RDS, DMS, EMR, ECS, Lambda, ElastiCache/Redis, SageMaker, Cribl e.t.c.
→ SQL, Hadoop, Docker, C#/.NET, Java, Shell Script, JavaScript, HTML/CSS.
→ Jenkins, JAMS, Bitbucket, MLOps, DevOps, Databricks, Domino Data Lab, Dataiku, e.t.c.

Experience

Lead Machine Learning Engineer – Finra, Rockville, MD ..... 2018.02 - Present

- Strategic planning for leveraging AI/ML to make cyber security more proactive, efficient and effective. Support Advanced Analytics strategy to ensure that models, data, tools and infrastructure are secured.
- Led the R&D Security Insights from VPC flow logs project, built infrastructures using Jenkins and Cribl for flow logs ETL, data preprocessing, feature engineering, model training and inferencing. Built an autoencoder model to detect abnormal traffic from over 20 applications' flow logs and a Power BI dashboard to demo findings for business users.
- Developed a Jenkins pipeline to containerize ML models and web applications, deployed to Production via AWS ECS.
- Led the R&D Funding Portal (FP) Risk Monitor System project to help the FP review team to monitor risk of the FPs' webpage contents. Built three deep learning models to detect potential Finra rules violations from the FP landing pages' contents using BERT, analyze sentiment of the investor comments on the FP discussion board, and compute webpage screenshots similarities between different review cycles to detect content changes using the image embedding vectors (hidden states of ResNet50).
- Led the R&D Hotmail - Corpus of Evidence project to help examiners to reduce noise among 1.22 Million email records. Conducted Network Traffic Analysis (NTA) and Community Detection to reduce the total 635K emails to 198K. Used Transferring Learning (BERT) to encode each email's content to a 768-dimensional semantic vector, and then applied dimension reduction (Truncated SVD) and clustering (K-Means, DBSCAN) to cluster the 198K emails into ~90 clusters. Conducted Topic Modeling on each cluster using Latent Derilicht Analysis (LDA) and Non-negative Matrix factorization (NMF) and ranking each cluster of emails based on the number of security symbols detected using Entity Extraction.
- Built and deployed a documentation classification model (NLP) with an 81% accuracy using CNN to help analysts from the department of Advertising Regulation to review securities related advertisements and communications submitted by Member Firms of Finra. It detects suspicious patterns among ~75k annual filings and automatically classifies those web advertisement documents to help analysts prioritize their work and mitigate the biases.
- Built an RNN model (LSTM/GRU) to classify Finra enforcement documents into 17 categories, achieving ~70% accuracy and using Python, PyTorch, NLTK, Gensim, Keras and scikit-learn on AWS G3 instance with CUDA Toolkit 9.2 and cuDNN 9.2. Tackled highly imbalanced dataset with down-sampling. Tuned the model parameters using grid search, and demoed the methodologies in front of the whole Member Regulation Technology Department of about 80 audiences.
- Won the Grand Prize and the 2nd place among the *It's Who You Know That Matters* group in the 2020 FINRA Createathon. Built a Graph Convolutional Neural Network (GCNN) model to help examiners to monitor the risks of Finra Member Firm FinOPs using PyTorch and DGL.
- Won People's Choice Award and the 1nd place among *the Amazing Spider Bot* group in the 2019 FINRA Createathon. Selected as a three-month R&D project mentioned above.
- Won People's Choice Award and the 2nd place among the Corpus of Evidence group in the 2018 FINRA Createathon. Applied LSTM/GRU model to classify regulatory coordinators' (RC) emails into different risk categories based on the content, conducted grid search on 400+ different combinations of hyperparameters and achieved an accuracy of 74%. Presented the results to the Finra Management Committee including CEO and CTO.

Big Data Engineer Consultant at Finra – ConsultNet LLC, Rockville, MD.....2017.01 – 2018.02

- Built a deep learning POC model (RNN/LSTM/GRU) to classify ~9000 customer complaints into 10 categories with 91% accuracy, using both Keras/TensorFlow and PyTorch. Provided mentorship to a summer intern.
- Won 1st place in the 2017 FINRA Createathon among the Unconventional Thinking group (17 teams in this group). Applied Random Forest and XGBoost e.t.c algorithms on broker-dealers' Financial and Operational Combined Uniform Single (FOCUS) Reports to re-evaluate their liquidity risk levels, achieved 60% accuracy overall and 75% accuracy in identifying high risk firms in liquidity. The results are validated by internal regulatory coordinators (RC) and received very positive feedback.
- Applied machine learning algorithms (regression, random forest) to build the High Risk Representatives (HRR) predictive model to detect suspicious behaviors among securities broker-dealers and investment advisors. Exacted and aggregated data from multiple databases for feature engineering using AWS EC2, EMR Cluster and S3. Used Lasso and random forest for feature selection.
- Built data ETL pipelines using JAMS and Jenkins for model development and ad-hoc analysis. Deploy and maintain models in the SDLC environment using both internal data management server and AWS services like EC2, S3, RDS, DMS.
- Completed the Firm Address Matching project using Python/fuzzywuzzy for business users to find out which broker-dealer (BD) firms were using virtual office (e.g., P.O Box) addresses to conduct their business. The results received very positive feedback from business users as they helped identify potential fraudulent behaviors by investment firms.
- Converted to full-time employee at Finra on 2018.02.

Sr. Predictive Modeler & Full-Stack Software Engineer – Health Integrity, Baltimore, MD 2014.06 – 2017.01

- Full-stack developer responsible for the design and development of the company's main product PLATO platform using .NET MVC framework with MS Visual Studio and SQL Management Studio. Handled monthly data ETL and ad-hoc data requests from clients.
- Major contributor of developing a sophisticated query engine "CLEAR System to System (S2S)" application, for data mining on federated public records data sources supported by Thomson Reuters Corp. Build complex data models and logic filters to target specific individuals and businesses and generate summary files and human-readable reports by parsing XML raw data stream.
- Applied both supervised and unsupervised ML algorithms to build the Atypical Antipsychotics Prescriber (ATP) model and Trio Prescriber model to detect patterns of fraud, waste or abuse (FWA) in Medicare Part D data. Worked closely with other statisticians and subject-matter experts (SMEs) to understand medical record data, capture business requirements, build and validate models to enhance and extend the FWA detection processes.
- Automated and optimized model monthly run programs and delivered monthly high risk pharmacy/prescriber reports to the clients. Those reports had led to 71 investigations opened, 35 of which were referred to law enforcement agencies and 11 have been accepted, as of May 2016.

Medical Informatics Analyst Intern, Wake Forest Baptist Health, Winston-Salem, NC 2013.06 - 2013.08

- Implemented a decision tree model for automated inference of patient problems from medical records.
- Generated a 10-page report of the models for further identification using SQL on Oracle Developer.

Data Analyst Intern, Wake Forest Baptist Health, Winston-Salem, NC ..... 2012.06 - 2012.08

- Analyzed the characterization of over 810 million clinic data records using Oracle SQL Developer.
- Performed distribution analysis and geographic study in Tableau, generated an 83-page report and delivered to hospital leaders and project directors.

Research Assistant, Wake Forest University, Winston-Salem, NC ..... 2011.11 - 2013.08

- Built statistical models using Markov chain to study DigitalAnts' random walks on different distributions of pheromone.
- Completed a thesis with two proved theorems and closed-form solutions for the optimal pheromone distribution on one-dimension grids.

Education

M.S. in Computer Science, Wake Forest University, full scholarship, GPA: 3.89/4.00	2013.08 - 2014.08
M.A. in Mathematics, Wake Forest University, full scholarship, GPA: 3.82/4.00	2011.08 - 2013.08
B.S. in Applied Mathematics, Wuhan University, China, GPA: 3.52/4.00	2007.09 - 2011.06