Dizhou (Bruce) Wu

310 E Springfield Ave, APT 210B · Champaign, IL 61820 · (217) 607-3789 · dizhouw2@illinois.edu

EDUCATION

University of Illinois at Urbana-Champaign Urbana, IL
Master of Science in Financial Engineering **GPA**: 3.8/4.0 Aug 2018-Dec 2019
Relevant Coursework: Deep Learning, Machine Learning, Financial Computing, Risk Management, Financial Derivatives

University of Liverpool Liverpool, UK
Bachelor of Science in Mathematics with Finance **GPA**: 3.7/4.0 First Class Honors (10%) June 2018

EXPERIENCE

CreditNinja Chicago, IL
Data Science Intern, Risk Sep 2019-Dec 2019
Domain Knowledge in Credit Risk Modeling Utilizing Advanced Data Analytics/Modeling

- Cleaned and analyzed a variety of data sources (Census Bureau, IRS, Data Vendor) to generate predictive feature engineering (e.g income-payday mapping, employment) to support the risk/modeling team.
- Performed time series analysis on bank account balance data to extract information on aggregated payment behaviors.
- Utilizing AWS SageMaker, Periscope data to perform all the data analysis/modeling in an agile development setting (Jira).

Calamos Investments Naperville, IL
Quantitative Analyst Intern (Applied Machine Learning) Jun 2019-Aug 2019
Proficiency in Feature Engineering, Feature Selection, Financial Machine Learning and NLP.

- Built a multiprocessing web-scraper to collect corporate filings (10-Ks, 10-Qs, 8-Ks) for all publicly traded companies from the SEC EDGAR database and performed extensive text data cleaning and preprocessing.
- Performed NLP analytics on the filings to generate 3 types of factors: Distance(Word2Vec Embedding based Word Mover Distance, YoY similarity), Complexity(log file size, readability index), Sentiment(Loughran-McDonald dictionary, Vader, Textblob polarity) with/without tfidf weighting and ran single factor analysis, multi-factor combination, and risk attribution using modified Alphalens.
- Implemented and optimized machine learning algorithms (MALTA(Season and Hedge), Time Series CV, SHAP Feature Importance, Meta-Labeling, Boruta Feature Selection) , deep learning models (CNN, GRU, LSTM, MultiHead-Attention Transformer), and factor orthogonization (Schmidt, Canonial, Symmetry) in the proprietary quant library for predictive analytics in production environment.
- Implemented Market Regime models using unsupervised learning algorithms like KMeans, DBSCAN/HDBSCAN, Agglomerative, Affinity Propagation, DeBaCI, Mean Shift clustering and FFT/Hidden Markov/Jenks natural breaks for univariate time series data.

CME Group Chicago, IL
Data Science Practicum Student (Anomaly Detection) Jan 2019-May 2019
Extensive experience on Building Models, Training and Testing, and Deploying Models in Tensorflow/Keras/Python.

- Analyzed corn and E-Mini S&P500 futures minute-level OHLCV data aiming to detect the market anomalies using statistical modeling (ARIMA, GARCH, Facebook Prophet) and deep learning (2-D CNN and RNN/LSTM with Autoencoders.).
- Validated anomalies detected by visualization (plotly) on the selected calendar events (FOMC, USDA) and the trader-defined events.
- Performed extensive feature engineering such as the realized variance (RVol) for improving training classification models including XGBoost, Random Forest, Bi-directional LSTM with Attention.
- Performed transfer learning for a large number of products within different sectors. Obtained improved performance over the baseline.
- Built Python GUI applications for human-labeling anomalies to generate robust labels as input for the news impact prediction model.

PROJECTS

Data Science and Machine Learning Project Highlights (Python, Tensorflow, Pytorch, SQL) Sep 2018-Sep 2019

- ***Deep Learning and Natural Language Processing****:*
  - Image to image translation using cycle GANS: reimplemented the ICCV 2017 paper to perform paired image translation.
  - Jigsaw Toxicity Classification: Fine-tuned pre-trained BERT model and evaluated AUC and background-subgroup AUC metrics.
  - Quora Insincere Question Classification: Tested and compared different embeddings (glove, fasttext, paragram) and their ensemble with deep learning architecture and techniques (Bi-CuDNNGRU, Capsule, Attention-LSTM, Cyclic LR)
  - Two Sigma News Prediction: Combined text features and OHLCV features with Scipy CSR Matrix to train a voting lightGBM, a stacked LSTM and a MultiHead-Attention models in Keras and Tensorflow.
- ***Fraud Detection and Credit Risk Forecasting:***
  - Credit Card Fraud Detection: Performed dimensionality reduction t-SNE, PCA, and truncated SVD. Applied imbalanced-sampling methods as random over/under sampling, SMOTE and Tomek for classification. An ensemble voting classifier consisting of Logistic Regression, SVM, Random Forest, and XGBoost/DART-XGBoost was trained to obtain a F1 score of 0.93.
  - Lending Club Loan Analysis: Evaluated Naïve Bayes classifier with up/down sampling on text features in the loan dataset. Explored the marginal predictive power increase of the mapping between Zip Codes and mean household income.

Portfolio Management and Quantitative Strategies (Equity & Fixed Income) (Python) Aug 2018-Mar 2019

- Applied PCA to risk factors model construction and tested 5 alpha factors(e.g Cross-Section/Time Series Momentum, Breakout) with evaluation metrics on Quantile, Turnover, and Sharpe Ratio and combined them using Random Forest for enhancing meta-alpha.
- Constructed an index-tracking portfolio and rebalanced optimal holding portfolio with constraints over time.
- Built a realistic backtester using the Barra data to perform portfolio optimization that includes transaction costs and conducted performance attribution to identify the major drivers of PnL

SKILLS

Programing: Python, SQL, H2O.ai, R, C/C++, MATLAB, MongoDB, Tableau, Bloomberg Terminal, Spark, AWS, Google Cloud Platform