Yang Lin Ph. D.

Email: linyangnyc@gmail.com Phone: (917)-930-7106
Linkedin: http://www.linkedin.com/in/yang-lin-ny Github: https://yangnyc1024.github.io/

SUMMARY

- 10+ years of **research experience** in the area including optimization, statistics, machine learning and risk
- 5+ years in **machine learning** (ML) models in technology, finance/insurance, IoT/ energy, and healthcare areas
- 3+ years in **project experience** on applied artificial intelligence (AI), data science, web development and mobile development
- Significant ability in communication, teaching, and work independently or as a part of a team

PROFESSIONAL EXPERIENCE

Research Scientist **Dec 2021 - Present**

Munich Re/ HSB

- Contributed to the world's first AI insurance product (*Insure AI*) team, offering performance guarantee solutions to **manage AI risks**, my focusing areas included financial fraud detection, IoT anomaly detection and AI cyber among others
- Led the risk assessment for evaluating AI product design, ML system performance and latest ML methodologies
- Led the development of statistical pricing models to offer insurance solutions for AI products
- Conducted research on distribution-free uncertainty in ML models via conformal prediction, aiming to deliver performance guarantees and improving risk assessment in machine learning algorithms
- Partnered with engineering teams to develop automotive risk assessment and pricing tools, leveraging python for backend development to evaluate ML algorithm performance on Databricks
- Worked alongside data science teams to implement **federated learning** and **deep learning** (LSTM) models for anomaly detection in sensor data, utilizing Azure as the platform.

Visiting Researcher /Adjunt Professor **Aug 2020 -Dec 2021**

Stevens Institute of Technology, Hoboken, NJ

- Worked on stochastic models and algorithms to quantify physical and financial risks of power systems at *Lei Wu's Energy Lab*
- Spearheaded and delivered an advanced course on ' MA 576 Optimization for Data Science' in the ***Data Science*** *Program*
- Designed and implement **reinforcement learning models**(approximate dynamic programming) for maximizing the value of pumped-storage hydroelectricity(PSH) resources in large scale electricity market and increased the profit by 11%
- Applied parallel computing to speed up the offline learning algorithm by 10 times to obtain the arbitrage policy by python and Gurobi

Doctorial Researcher **Sep 2014 -Aug 2020**

Stevens Institute of Technology, Hoboken, NJ

- Researched on statistical inference of stochastic optimization problems in quantitative risk management and machine learning
- Created kernel smooth model by nonparametric **statistical methods** to approximate risk-neutral/averse composite stochastic optimization problems
- Established the consistency, convergence, stability, and bias reduction of the smooth estimators and extended the method to regression models, classification problems, and optimization problems with financial risk metric
- Built smooth data-driven **optimization model** and reduced mean square error(MSE) of on the average value at risk(AVaR) model by 16.7% and higher-order measure of risk by 25%

SKILLS

Programming Languages: Python, Java, JavaScript, SQL, Go, C++, Matlab, R
AI Fundamentals: Machine Learning, Deep Learning, Reinforcement Learning, Statistics and Optimization, etc
AI Applications: Financial Fraud Detection, IoT anomaly detection, Recommendation Systems, CV, NLP, AI cyber, etc
Analytics: A/B Testing, ANOVA test, DS Pipeline (cleaning, EDA, interpretation, model)
Big Data Engineering/ Database: MySQL, Spark, ETL
Framework and Platforms: Spring, Hibernate, React, AWS, Databricks, Microsoft Azure, GCP
Web Development: Node.js, HTML & CSS, React

EDUCATION

Ph. D. in Applied Mathematics May 2020
Stevens Institute of Technology, *Hoboken, NJ*
Dissertation: Kernel Smoothing in Sample-Based Optimization; Advisor:Darinka Dentcheva

M.S. in Mathematics May 2014
New York University, *New York, NY*

B.S. in Mathematics, Information and Computational Sciences May 2011
Xiamen University, *Fujian, China*

PROJECT EXPERIENCE

Topic includes business index prediction, recommendation engine system, semantic analysis, image recognition and website develop

Movie Recommendation Engine Development

- Developed data ETL(extract, transform, load) pipeline to analyze movie rating dataset with Spark SQL
- Implemented the Alternative Least Square algorithm, tuned hyper-parameters with Spark ML toolbox, and monitored data processing performance via Spark UI on AWS(RMSE=0.7)
- Provided personalized movie recommendations and developed user-based approaches to handle system cold-start problems

Semantic Analysis for Youtube User Comments Dataset

- Built a machine learning model to analyze and classify users based on their Youtube videos comments on cats and dogs
- Extracted and transformed users' comments via RegexTokenizer and Word2Vec with SparkML pipeline
- Trained logistic regression, Random Forest, and GBT models, tuned hyperparameters, and selected the best model(AUC=0.91)
- Extract insights about cat and dog owners and identified creators with cat and dog owners in the audience

Natural Language Processing(NLP) and Topic Modeling on User Review Dataset

- Discovered and studied main topics of an E-Commerce business customer review dataset for watch by NLP
- Processed review texts to extract features by tokenization and stemming
- Constructed the term frequency-inverse document frequency(TFIDF) matrix to build the unsupervised learning model
- Trained model including K-Means and Latent Dirichlet Allocation(LDA) to extract the main topics on watch review(good, nice, not cheap, etc.)

Customer Churn Prediction in Finance Industry

- Built algorithms for the bank to predict customer leaving based on information and performance records
- Trained supervised learning models including Naive Bayes, Logistic Regression, Random Forest(RF), Gradient Boosting tree(GBT), K-Nearest Neighbors, applied regularization with optimal parameters to overcome overfitting
- Evaluated model performance (Accuracy 0.86) via k-fold cross-validation technique and analyzed feature importance to identify the top factors that influenced the results

Car Image Recognition and Classification based on Deep Learning

- Built a deep learning model to classify the car image
- Trained Convolutional Neural Networks (CNN) model by utilizing the pre-trained model and fine-tune on current dataset
- Deployed the built transfer learning model and predicted the car's name on single images (Accuracy=0.95)

Stock Prices and Market Index Prediction

- Analyzed the volatility feature of stock price and market index based on Pytorch
- Built a deep learning time series model (Long Term Short Memory) by 7 days in advance to predict the stock price
- Trained LSTM model by changing activation and regularization function on GPU
- Deployed the built LSTM model as a service to predict the variation of the S&P 500 index

Video+: A Personalized Twitch Resources Recommendation Engine

Front End:

- Designed and built a full-stack web application for users to search twitch resources(stream/video/clip) and get recommendations.
- Built a web page with rich + user friendly experience using React and Ant Design

Back End:

- Created three Java servlets with RESTful APIs to handle HTTP requests and responses.
- Used MySQL database on Amazon RDS to store position data fetched from Twitch API
- Designed algorithms (e.g., content-based recommendation) to implement twitch resources recommendation.
- Support login/logout and favorite collection and deployed to Amazon EC2 for more visibility

SELECT PUBLICATIONS

- Sukrita Singh, Neeraj Sarna, Yuanyuan Li, Yang Lin, Agni Orfanoudaki, Michael Berger. **Distribution-free risk assessment of regression-based machine learning algorithms**; *under review by NeurIPS, 2023*
- Yikui Liu, Yang Lin, Lei Wu, etc. Optimal SOC Headroom for Pump Storage Hydropower Dispatched by ISOs: Maximizing Revenue in Day-ahead and Real-time Markets; *under review by IEEE Transactions on Sustainable Energy, 2023*
- Darinka Dentcheva, Yang Lin, Spiridon Penev. **Stability and Sample-Based Approximations of Composite Stochastic Optimization Problems**; *Operations Research, 2022*
- Darinka Dentcheva, Yang Lin. Bias Reduction in Sample-Based Optimization; *SIAM Journal on Optimization, 2022*
- Yonghong Chen, Lei Wu, Ross Baldick, Yang Lin, Yikui Liu, etc, Modeling and Optimizing Pumped Storage in a Multi-stage Large Scale Electricity Market under Portfolio Evolution; *Missouri University of Science and Technology*, 2022