Annan Chen

Phone: (646) 258-5178 | **Email**: ac4619@columbia.edu | LinkedIn | GitHub | Website

Technically sophisticated with talent in **Programming** and **Descriptive/Predictive/Prescriptive analytics** utilizing a range of **Statistical** and **Machine Learning** methods. **Strong academic credentials** and **leadership** skills with experience gathering information, testing hypotheses, and developing recommendations. Fluent in English with ability to articulate complex ideas and **communicate effectively with audiences** across all levels of technical understanding and organizational position.

EDUCATION

Columbia University New York**,** NY
M.S. in Operations Research (STEM, Machine Learning/Data Analytics track), GPA: 3.95/4.0 Sep 2019 – Dec 2020
Courses: Cloud Analytics (Scala, Spark), Machine Learning (Python), Deep Learning (Python), Database (SQL), Data Analytics (Python), Algorithm Trading (Python), Simulation (Python), Tools for Analytics (Python, SQL, Linux, Django), Business Analytics (R)

The Chinese University of Hong Kong Shenzhen, CN
B.B.A. in Global Supply Chain and Logistics Management, First Class Honor, GPA: 3.81/4.0 (Rank: 2/300) Sep 2015 – Jun 2019
Honors: Presidential Award for Outstanding Student (1%), 2016-2019 Dean's List, 2016-2019 Academic Scholarship (top1%)

PROFESSIONAL EXPERIENCE

Flexstone Partners | *Private Equity Firm affiliated with Natixis* New York, NY
Research Intern Feb 2020 – Dec 2020

- Built machine learning and deep learning models (TensorFlow & Keras) to predict industry ETF Return/Sharpe Ratio/Risk/Calmar Ratio in the future 2-5 years and rank sectors respectively; achieved 85.7% test set Kendall's tau
- Models deployed in the research: Ridge, Lasso, SVR, XG Boost, DNN, LSTM, CNN-LSTM, GRU, ESN (Echo State Network)
- Extracted different macroeconomic indicators for each particular industry and utilizing RFE to select most important features
- Performed stochastic interpolation and Variance Gamma Process to transform quarterly data to daily time series

BrainCo | *Brain-Computer Interface (BCI) Firm incubated in the Harvard Innovation Lab* Boston, MA
Data Scientist Intern Sep 2020 – Dec 2020

- Developed machine learning and deep learning models via Python based on EEG (electroencephalography) data
- Performed signal processing and feature extraction via PCA, LDA, FFT, etc., and built Time-Series Classification algorithms

Blue Elephant Capital | *First Education-focused VC in China, Backed by TAL Edu Group (NYSE: TAL)* Beijing, CN
Project-Based Summer Associate (Remote) Apr 2020 – Aug 2020

- Improved 80% efficiency by automating business proposals (BP) collecting procedure via Python (Selenium), Google Form & Jira

The Coca-Cola Company Shanghai, CN
Data Analytics Intern for Strategy, Knowledge & Insight Department Jun 2019 – Aug 2019

- Forecasted beverage sales and volume in future 8 quarters; attained less than 3% MAPE using ARIMAX model in R
- Collected 10,000+ lines of beverage price data from e-commerce platforms (JD and TMALL) using Python (BeautifulSoup)
- Visualized and analyzed package and juice flavor data using Power BI and Tableau & provided business insights for R&D

Tencent Holding Ltd. Shenzhen, CN
Business Analytics Intern for Strategy Department within Interactive Entertainment Group (IEG) Apr 2018 – Aug 2018

- Facilitated 20 ml. downloads with 15 ml. monthly revenue (per game) achieved with new launching and publishing strategy during summer & led the cloud gaming research project (built framework for new cloud gaming platform START)

PROJECT EXPERIENCE

Morgan Stanley NLP Competition, Twitter Sentiment Analysis (Top 3 / 50 Teams), New York, NY Jun 2020 – Aug 2020

- Derived insights on Wealth Management industry by analyzing Twitter data with machine learning approaches using Python
- Extracted data related to Wealth Management topics, themes and service models using Tweepy and preprocessed data with NLTK
- Accomplished 96% top-2 accuracy (78% overall accuracy) on self-labeled dataset with fine-tuned Ensemble BERT model via Keras
- Built visualization dashboard of popularity and sentiment trend on Wealth Management firms, topics and services using Python

ML Prediction for 2020 U.S. Presidential Election, Columbia University, New York, NY Sep 2019 – Dec 2019

- Constructed machine learning models using Python to predict the 2020 U.S. Presidential Election result with US Census data
- Achieved 93% sign correctness in the validation set using Lasso in vote differences and visualized result with matplotlib
- Established ML models with Support Vector Regression, Gradient Boosted Tree, XGBoost, Lasso/Ridge Regression in Python

NBA Relational Database Web Application, Columbia University, New York, NY Jan 2020 – May 2020

- Designed entity-relationship diagram and corresponding SQL schema; entities included teams, players, stadium, game info, etc.
- Implemented relations and schema over PostgreSQL database server on Google Cloud Platform
- Developed Flask web application using Python, SQL, HTML with functions allowing users to perform search functions

SKILLS

Programming: Python (Pandas, NumPy, TensorFlow/Keras, scikit-learn, StatsModels, XGBoost, SparkMLlib, Spark SQL, nltk, matplotlib), R, SQL, Scala, Apache Spark, Git, Linux, Django, PyQt, Google Cloud, Tableau, Power BI, MATLAB, Adv. Excel
Analytics: Text Mining, Natural Language Processing, Sequential Model and Forecast, A/B Testing, Web Scraping, Cloud Analytics