

## Wes Shi

781-718-8038 | [wesshi818@gmail.com](mailto:wesshi818@gmail.com) | [Portfolio](#) | [LinkedIn](#)

### EDUCATION

Columbia University New York, NY

Sep 2022-Dec 2023

MSc Operations Research (STEM) - **Analytics** track

GPA 3.93/4.0

Brandeis University Waltham, MA

Jan 2019-May 2022

BSc Applied Mathematics & Quantitative Economics

GPA 4.0/4.0

### SKILLS

**Programming:** Python (Sklarn, TensorFlow, Beautiful Soup, NLTK, NumPy, Pandas, Matplotlib, Plotly), **SQL**, R, Java

**Methods:** A/B Testing, Time Series Forecasting, Natural Language Processing, Statistical Modeling, Web Scraping

**Tools:** AWS S3, Databricks, Excel, Hadoop, Word, PowerPoint, Spark, GitHub, Git, Redshift, Stream lit, Snowflake, Tableau

### PROFESSIONAL EXPERIENCE

#### System2

Jun 2023-Aug 2023

Data Scientist Intern (Financial Consulting)

New York, NY

- **ETL:** Governed and optimized ETL process of actively trading stocks' **alternative data (1 TB+)** using **AWS S3** and redshift database, reduced data extraction time by **60%+**
- **Machine Learning:** Performed quantitative portfolio analytics based on market risk data, adapted **random forest & PCA** for selecting **6+** influential risk metrics, supported making investment decisions based on selected technical indicators
- **Data Visualization:** Developed **10+** customizable app (Stream lit) with scalable **Python & SQL** pipeline for visualizing portfolio return prediction based on **LSTM** and **XGBoost**, achieved MSE of 1.7 on short-term prediction
- **A/B Test:** Designed and analyzed A/B test for clients, evaluated the effectiveness of new credit card offer tag by **T-test**, identified **~3% growth** on active users
- **Market Research:** Constructed pipeline for scraping **8k+** gyms' info and automatically fulfill/update redshift database; innovated **market growth estimation algorithm** based on **geospatial data** and provided business intelligence strategy

#### Align Partners

Jun 2021-Jun 2023

Business Intelligence Engineer (Wealth Management)

Remote

- **Machine Learning:** Deployed agglomerative clustering model on Databricks (AWS), applied elbow method to optimize clustering, increased asset allocation recommendation efficiency by **40%+** through effective customer segmentation
- **Business Intelligence:** Developed **20+ tableau dashboards** to visualize financial plans and investment return for customers

#### DG Venture Law Firm

Jan 2023-May 2023

Product Manager Intern (Legal Service)

New York, NY

- **Product Design:** Collected business needs of NDA review product partnered with legal department, identified **5+ end-user requirements** (e.g., cloud-based pipeline), developed detailed documentation
- **Product Development:** Led a team to design and develop an **end-to-end solution** for detecting, highlighting clauses containing deal-breaking information in NDAs, providing modification suggestions
- **Text Classification:** Implemented text feature extraction with bag of words model and trained classifiers including LinearSVC (f1 ~82.3%), Multinomial Naïve Bayes for clause classification
- **NLP Modelling:** Fine-tuned **BERT** model and **neural network** to perform clauses' classification, achieved 95%+ f1 score; utilized corresponding deal-breaking info searching, captured **99%+ improper clauses**, saved **20+** minutes per NDA review

#### JOANN

May 2022-Aug 2022

Data Science Intern (Retailing)

Hudson, OH

- **Exploratory Data Analysis:** Performed EDA to raw supply chain data from HANNA using Python & SQL; identified 14+ types of inaccurate records and automated data cleaning pipeline, increased **data integrity by 10%+**
- **Feature Engineering:** Constructed **SQL** table including 280+ ship routes' distance, origin, and destination ports for revamping the original geo-map straight distance, improved feature accuracy by **52%**
- **Machine Learning:** Cooperated with analytics team, applied advanced analytical methods including XGBoost in cargo transit time prediction, achieved model error reduction from 14 to 8.4 days, **improved warehouse efficiency by 30%+**
- **Business Intelligence:** Analyzed the effect of replacing HAP with snowflake, including reducing the query time consuming by **60%+**, presented the insights to stakeholders through **Power BI dashboards**

### PROJECT

LLM Research Assistant (Columbia University)

Sep 2023-now

- **Literature Review:** Collaborated with fellow RA on literature review for LLMs in social dilemmas, examining their dynamics in 5+ traditional games, including prisoners' dilemma, ultimatum game, etc.
- **LLM application:** Pioneered prompt engineering for LLMs in sequential multi-agent social dilemmas like fruit gathering; developed **test framework and pipeline** to display real-time actions of different LLMs based agents