JAI AGRAWAL

agrawaljai2000@gmail.com | www.jaiagrawal.com | (213) 268-9080

EDUCATION

University of Southern California (United States) Anticipated: May 2024
Master of Science in Applied Data Science (5-year accelerated degree)
Graduate Coursework: Advanced Machine Learning, Applied Natural Language Processing, Foundations of Data Mining, Data Science at Scale & Analysis of Algorithms

Bachelor of Science in Economics and Data Science
Undergraduate Coursework: Probability Theory, Applied Differential Equations, Linear Algebra, Economics Applications of Machine Learning & Economic Consulting and Applied Managerial Economics

Teaching Assistant: Introduction to Data Analytics, Programming in Python

GPA: 3.73 (Undergraduate), 3.75 (Graduate)
Dean's List: Fall 2019, Spring 2020, Fall 2021, Spring 2022, Fall 2022, Spring 2023, Fall 2023

WORK EXPERIENCE

Data Scientist Intern, Roark Capital & Miller's Ale House Jun-Aug 2023
Roark Capital is a leading PE firm focused on franchised business investments such as the U.S.-based dining chain Miller's
- Developed a Palantir-based analytics program utilizing the Apriori algorithm, which enhanced CRM data precision and identified key product pairings, driving a 12% increase in cross-sales through promotional campaigns
- Applied the FP-Growth algorithm to discern frequent item-sets, streamlining inventory management which led to a 18% reduction in excess stock and a further decrease in related costs

Machine Learning Intern, Vincilium – Distributed Ledger Tech May-Jul 2022
Vincilium is a DLT-based software development company operating in New Jersey
- Engineered a Python-based NLP chatbot leveraging Keras, NLTK, and TensorFlow, which empowered users to conduct advanced data retrieval from a PostgreSQL database, enhancing user experience and operational efficiency
- Developed a REST API that handled GET and POST requests, optimizing application data exchange and significantly improving backend system interactivity
- Contributed to the application's AI capabilities, resulting in a 25% increase in query resolution speed and a 30% reduction in manual data extraction tasks

PROJECTS

Quantitative Analysis on Canada Post Pricing Using Price Elasticities in Python 2023
- Conducted comprehensive regression analysis on Canada Post's admail pricing using Python, achieving a high R-squared value, which revealed significant seasonal trends affecting profitability.
- Determined admail demand elasticity at -0.42 through cross-price elasticity analysis, suggesting room for strategic price adjustments with minimal demand disruption.
- Recommended data-driven pricing strategies based on market analysis, positioning Canada Post to capitalize on revenue-generating opportunities.

Analysing Eminem's Lyrics Using TF-IDF Topic Modelling in Python 2021
- Designed a TF-IDF topic modeling system using Python and the NLTK library, extracting thematic patterns from Eminem's discography for linguistic analysis.
- Automated data extraction from Genius API, preprocessing lyrics with NLP techniques to feed into the topic modeling algorithm, revealing distinct thematic clusters.
- Identified opportunities for nuanced lyrical analysis across albums, setting a foundation for future work in comparative text analysis and cultural studies.

LEADERSHIP

Vice-President, Expat Society (USC) Aug 2022 - Present
- Spearheaded the design of a JavaScript web-app for the Expat Society, fostering community engagement with event hosting and real-time communication features, achieving a 30% member adoption rate
- Catalyzed a 50% increase in club membership by leading collaborative efforts, demonstrating effective team-building and strategic outreach.

TECHNICAL SKILLS

Languages & Tools: Python, SQL, R | Excel, Tableau, Spark, MongoDB, Git, Foundry (Palantir)
Libraries: Pandas, NumPy, PySpark, Seaborn, XGBoost, BeautifulSoup, Sklearn, PyTorch, TensorFlow, NLTK
Statistical Modelling: Hypothesis Testing, Regression Analysis, Instrumental Variables, Time Series Analysis
Machine Learning: Supervised (Linear/Logistic Regression, Decision Trees, SVM, KNN), Unsupervised (Content-Based Filtering, K-Means, Hierarchical Clustering) Dimensionality Reduction (PCA), Resampling (Bootstrap, Cross-Validation)