Harold Ng

Email: haroldng2001@gmail.com | **Phone:** (925) 214-8620 | **LinkedIn:** linkedin.com/in/harold-ng-986a5b252

EDUCATION

University of California, Irvine **Dec 2024**
Master of Data Science **GPA**: 3.92
University of California, Los Angeles **Jun 2023**
B.S. in Applied Mathematics w/ Specialization in Computing **GPA:** 3.98

TECHNICAL SKILLS

Programming Languages: Python, SQL, R, C++, Javascript, HTML, CSS3, MATLAB
Tools/Frameworks/Software: TensorFlow, PyTorch, Scikit-Learn, Matplotlib, Pandas, Numpy, PostgreSQL, MongoDB, Tableau, Snowflake, Jupyter Notebooks, Cassandra, Neo4j, Apache Spark, Apache Flink

EXPERIENCE

CooperVision *San Ramon, CA* **Jun 2024 - Nov 2024**
Data Analytics Intern

- Developed a **Fourier Transform-based model** in Python using NumPy and Pandas to remove seasonality from **523,000+ order records**, improving trend analysis and enabling real-time performance alerts for 115+ sales representatives, resulting in a **26.3% increase** in operational efficiency.
- Built an **automated data cleaning pipeline** in Python, generating datasets that revealed true demand patterns and **improved forecasting accuracy by 19.7%**, enabling better decision-making across 5 departments.
- Conducted **sentiment analysis** on 10,324 customer feedback entries using **Python and NLP techniques**, producing actionable insights that drove a **14.8% improvement** in customer satisfaction.
- Utilized SQL and Tableau to clean, analyze, and visualize datasets with over **1.2 million rows**, improving data accessibility and supporting data-driven forecasting accuracy for cross-functional teams.

RESEARCH PROJECTS

Sentiment-Enhanced Recipe Scoring System **Mar 2024**
Project Lead

- Led a team to develop a sentiment-enhanced recipe scoring system, leveraging a dataset of 17,500+ recipe reviews to predict algorithmic weight assigned to user comments, **improving recommendation relevance by 13.7%**.
- Performed **sentiment analysis** using VADER and TextBlob libraries, computing polarity and subjectivity scores for **50,000+ sentiment attributes** to refine user preference insights.
- Implemented and optimized **Multilayer Perceptron (MLP) Regressor** and **Gradient Boosting Regressor models**, achieving an **18.4% reduction in mean squared error** compared to baseline models.

Microsoft Single-Source Knowledge Base (SSKB) Transition Analysis **Dec 2024**
Project Lead

- Led a team of 6 to evaluate the impact of Microsoft's transition to a single-source knowledge base (SSKB) for internal customer support content, analyzing **60,000+ support case records** and Evergreen article data to assess operational efficiency and customer satisfaction.
- Developed statistical models, including Generalized Linear Models (GLM) with log-link functions, to analyze investigation time and clicks, revealing a **12% reduction in investigation time** and a **65% decrease in investigation clicks** post-SSKB implementation, indicating significant improvements in workflow efficiency.
- Presented findings to **Microsoft stakeholders**, highlighting the SSKB's role in reducing operational inefficiencies and improving customer satisfaction, with recommendations for further refinement of the knowledge base.