

ENSUN PAK

341-209-0236 | ensun.pak.dev@gmail.com | Palo Alto, CA | [linkedin.com/in/ensunpak](https://www.linkedin.com/in/ensunpak) | github.com/ensunpak

EXECUTIVE SUMMARY

- Experienced data specialist with over 10 years in data analysis and data science.
- Expertise in leveraging data analytics, machine learning, predictive modeling, and statistical inference to drive business growth and optimize operations.
- Proven track record in managing high-impact projects and leading teams in fast-paced environments.
- Skilled in developing scalable data solutions, conducting in-depth analyses, and implementing innovative strategies to enhance decision-making processes.
- Proficient in Python, Pandas, scikit-learn, XGBoost, SQL and Google Looker.
- Adept at collaborating with cross-functional teams to achieve organizational goals and improve customer satisfaction.

EXPERIENCE

Data Scientist

Four Analytics, Inc.

Palo Alto, CA (Remote)

September 2023 - Present

- Engineered infrastructure to enhance XGBoost classifier models, achieving a 61% accuracy in sports win probability predictions. Leveraged Python, Hyperopt, and grid search optimization techniques.
- Validated outputs of XGBoost classifier models by conducting Monte Carlo simulation on model performance over time and used the Wilcoxon signed test to determine the statistical significance of the results.
- Automated daily data scraping from baseball-reference.com using Python, SQLite, and cron on Google Cloud infrastructure, ensuring timely and accurate data availability.
- Designed and implemented interactive dashboards in Google Looker, providing insightful visualization of key performance indicators sourced from Pregame.com data, and facilitating informed decision-making processes.

Senior Data Analyst

Grab Singapore

Singapore

July 2016 - June 2022

- Developed predictive models using fleet operations and ride-share transactional data to assess business expansion feasibility in new Indonesian cities, resulting in a 24% revenue increase post-implementation.
- Collaborated with the Sales team to employ data analysis techniques in identifying high-quality customers from a pool of 70,000, achieving a low payment default rate of 0.6% in the business portfolio.
- Led a high-performing team in a fast-paced startup environment, surpassing productivity targets and maintaining an annual employee satisfaction rate of over 85%.

Credit Risk Data Analyst

OCBC Bank

Singapore

June 2012 - July 2016

- Led a team and collaborated with Sales and Product teams to reduce credit risk by 3% in a S\$950 million national home mortgage portfolio through quantitative data analysis using SAS.
- Developed and executed A/B testing strategies with SAS and FICO Blaze Advisor, resulting in a 9% improvement in customer sign-up satisfaction.
- Implemented data transformation processes, accelerating quarterly credit risk board reviews by 20% for faster report generation.

Credit Risk Data Analyst

Citibank

Singapore

2010 - 2012

- Recommended credit policies for the new Auto Loan exotic car segment using SAS, leading to a 34% increase in net revenue.
- Influenced successful credit policy tightening on the national Mortgage Loan portfolio through data analysis, resulting in a 0.9% reduction in monthly credit loss.

EDUCATION

Master of Science in Data Science, University of San Francisco

San Francisco, California

Bachelor of Arts in Information Systems, Queensland University of Technology

Brisbane, Queensland, Australia

PROJECTS

Slack-Based Request Message Parser Utilizing LLM

Developed Slack Bot using LLama3 LLM to categorize unstructured requests, reducing processing time from 5 minutes to 10 seconds. Tech: Ollama, Slack Bolt, LangChain, Python.

OCR-Based Construction Plan Data Extractor

Implemented OCR-based PDF extractor for construction plans, reducing data extraction time from 10 minutes to 30 seconds with 70% accuracy. Tech: pytesseract, pymupdf, Python.

ML-Driven Sports Investment Optimizer

Implemented daily ML pipeline using XGBoost and Hyperopt for sports league classifiers. Achieved 65% average model accuracy, contributing to 9% lift in investment performance (52% to 57%). Tech: Python, XGBoost, Hyperopt.

ML-Driven Janitorial Clean Cost Predictor

Implemented K-means clustering and XGBoost to predict janitorial clean job costs. Achieved RMSE of \$230 (6% error) compared to project manager estimates. Features: site size, retail type, city. Tech: Python, scikit-learn, XGBoost.

Kaggle Medical Prediction Competition - 2nd Place

Achieved 2nd place out of 43 teams in Kaggle medical indicator prediction competition. Developed Linear SVR model with 3.756 MAE, 0.011 behind 1st place. Tech: Python, scikit-learn.

A/B Testing for Media Streaming Optimization

Optimized browsing time for simulated media streaming provider using A/B testing. Achieved within 5% of optimal solution through grid-search and interaction testing. Constrained to 30 experiment units. Tech: Python, statistical analysis tools.