Bill Freeman, PhD

BillFreeman44@yahoo.com | https://github.com/billfreeman44 | https://www.linkedin.com/in/billfreeman44

Gaming Data Science Experience

Lead Data Scientist / Co-Founder at PureSkill.gg (Jan 2020 – Jun 2022)

- Founded and scaled an automated coaching platform that detects gameplay mistakes and directs players to targeted practice to improve specific weaknesses.
- Built and calibrated LightGBM models estimating win probability 200k times per match from game-state telemetry, enabling quantitative measurement of player decisions and actions.
- Trained model combining HDBSCAN, linear regression, and player rank to detect whether smoke grenades were correctly or incorrectly thrown per player per match.
- Used k-means clustering on player movement telemetry and weapon fire events to detect whether players were correctly counter-strafing during matches.
- Prioritized coaching recommendations using a scoring system based on win impact, mistake frequency, and difficulty to correct, adjusted for player rank so the system focuses on weaknesses specific to each skill level.
- Hired human coaches to provide manual ratings that would calibrate feedback for our customers.
- Processed 1.5M matches for 15,000+ players, including 250+ paying users.

Research Contributions and Other Projects

- Released public data on AWS enabling research for 50+ confirmed download requests.
- Supported the data with data docs, open source software, and tutorials enabling quick onboarding.
- Provided data and modeling advice leading to publication of a paper researching the differences in feature importance of a win-rate model between professionals, high skill players, and noobs.
- Advised masters student on converting complicated player telemetry data into a more understandable vector space using out-of-the-box algorithms such as Word2Vec. This methodology is especially applicable to games with non-Euclidean 3D space (ex: Valorant)

Industry Data Science Experience

Senior Data Scientist - Pricing Team, Kohl's (Jan 2025 - present)

- Led delivery of a deep-learning pricing system (Temporal Fusion Transformer, TFT), aligning data science and engineering teams to drive $50M in annual revenue through optimized discounting.
- Authored and executed the technical roadmap for the forecasting data science team of 6, establishing coding standards, experiment tracking, and review processes that enabled all team members to ship models safely to production.
- Acted as the primary business partner and final decision owner for pricing model launches.
- Mentored a summer intern, guiding feature imputation work that improved the accuracy of the production forecasting system.

Data Scientist - Search/Digital Signs Teams, Kohl's (Jun 2022 – Dec 2024)

- Owned personalization modeling for search, delivering Logistic Regression models that increased annual revenue by $2M, measured through A/B testing.
- Drove the adoption of embedding-based retrieval, leading development of product and query vectors that enabled semantic search across millions of SKUs.
- Refactored Scala pipelines to Spark, reducing code size 10x and cutting runtime in half.

- Delivered an LLM-powered classifier for customer feedback that reduced incident response time from days to hours by surfacing critical issues automatically.
- Solely built and rolled out LLM-generated marketing content to 500,000 in-store digital signs, establishing prompt frameworks, guardrails, and automated quality evaluation (LLM-as-a-judge).
- Partnered with Legal and AI Governance to design FTC-compliant validation workflows, enabling safe, company-wide rollout of generative AI content.

Data Scientist Contractor - RhoImpact.com (Mar 2019 – Dec 2019)

- Achieved 90% accuracy with TF-IDF and logistic regression classifying 100k medical docs daily.
- Developed investor-company recommender using collaborative filtering on past investments.

Data Scientist Contractor - Lowe's (Aug 2018 – Mar 2019)

- Built a greedy assortment optimization algorithm increasing sales while reducing waste.
- Built a data engineering pipeline with Hadoop/SQL to estimate stock levels from messy retail data.

Researcher - University of California, Riverside (May 2012 – Aug 2017)

- Published first-author peer-reviewed paper on galactic outflows at high redshift.
- Designed open-source Markov Chain Monte Carlo software to decompose galactic signals into a broadened component (indicative of outflows) and a narrow component (inherent emission).
- Took data using a $300,000,000 telescope for 15+ nights following a meticulous observing plan.
- Developed software to extract usable data from raw astronomical images for the largest survey ever conducted using the Keck Telescope; trained 10+ team members in using it.

Education

PhD Physics from University of California, Riverside - Riverside, CA

MS Physics from University of California, Riverside - Riverside, CA

BS Physics from Louisiana State University - Baton Rouge, LA

Highlighted Skills

Machine Learning: time-series forecasting, LightGBM, Logistic Regression, ensemble models, clustering (HDBSCAN), Temporal Fusion Transformers, embeddings, NLP, classification and regression, A/B testing, statistical modeling, feature engineering, recommender systems, personalization systems

Languages & Platforms: Python (expert), SQL (advanced),TensorFlow, Spark, Scala, distributed computing, production ML systems, AWS, GCP

Data Systems: Feature pipelines, large-scale telemetry, real-time and batch inference, experiment platforms, model deployment, ML architecture, data pipelines, simulations

Generative AI: Prompt engineering, LLM-as-a-Judge, content quality and safety evaluation, Gemini, ChatGPT, enterprise GenAI deployment

Publications

Analyzing the Differences between Professional and Amateur Esports through Win Probability. Xenopoulos, Freeman, and Silva, 2022

*The MOSDEF Survey: Broad Emission Lines at Z = 1.4 − 3.8**. Freeman et al. 2018