

SUMMARY

Data Scientist with a **Ph.D. in Physics** and 3+ years of applied experience building and deploying machine learning models, designing experiments, and translating large-scale datasets into measurable business outcomes. Proficient in Python (scikit-learn, TensorFlow), SQL, statistical modeling and end-to-end ML pipelines including supervised/unsupervised learning, causal inference, A/B testing, and anomaly detection across pricing, healthcare, and materials science domains. Proven ability to collaborate cross-functionally with stakeholders to deliver data science solutions that drive business impact.

SKILLS

Programming Languages: Python, SQL

Tools & Libraries: NumPy, Pandas, scikit-learn, TensorFlow, PyTorch, Spark, Hadoop, Git/GitHub

Data Science Techniques: Exploratory Data Analysis (EDA), Data Preprocessing, Data Cleaning, Train/Test Split, Hyperparameter Tuning

Data Visualization: Tableau, Matplotlib, Seaborn, Data Storytelling

Machine Learning: Supervised & Unsupervised Learning, Deep Learning, NLP, Classification, Regression, Feature Engineering, Predictive Modeling

Reinforcement Learning: Q-Learning, Deep Q Networks (DQN), OpenAI Gym, Stable-Baselines3

Data Analysis & Modeling: Statistical Analysis, A/B Testing, Causal Inference, Experiment Design, Hypothesis Testing, Model Evaluation and Development

EXPERIENCE

Data Scientist Fellow | Springboard | Remote | USA

May 2025 – Present

- Engineered end-to-end ML pipelines for 4 predictive analytics projects — pricing optimization, healthcare risk, and materials science — on datasets of 1,000–2,000+ records delivering measurable business impact across each domain.
- Optimized Random Forest, XGBoost, SVR, Logistic Regression, and KNN models via k-fold cross-validation and systematic hyperparameter tuning, achieving up to 17 percentage point accuracy gains (67% → 84%).
- Drove feature engineering and SelectKBest-based feature selection across all projects, reducing multicollinearity and strengthening model generalization on held-out test sets.
- Designed and ran A/B test frameworks and hypothesis tests to validate model improvements and support evidence-based decision-making across 4 domains.
- Communicated findings to cross-functional stakeholders through structured reports and Tableau/Seaborn visualizations, translating complex model outputs into clear business recommendations.

Tech stack: Python, scikit-learn, XGBoost, SQL, Pandas, NumPy, Matplotlib, Seaborn, Tableau, Git

Assistant Professor (Research Cell) | Poornima University, Jaipur | India

Aug 2023 – Jan 2025

- Led 3 data-driven thermoelectric thin-film research projects applying statistical modeling and experimental design; produced 2 peer-reviewed publications and 2 national conference presentations.
- Delivered 20+ technical sessions on data acquisition, EDA, and statistical analysis to cohorts of 30+ students; boosted lab data interpretation accuracy by ~20% across mentored cohorts.
- Mentored 100+ undergraduate students in experimental characterization and data interpretation, collaborating cross-functionally with faculty to reduce error rates across student projects.

Tech stack: Python, Statistical Analysis, Matplotlib, Experimental Design

Research Fellow | Department of Physics, MNIT Jaipur | India

Jan 2018 – Dec 2022

- Pioneered novel data analysis pipelines to evaluate nanomaterial performance and executed over 50+ experiments, raising thermoelectric figure of merit (ZT) by 12% validated across 3 independent operating conditions.
- Refined 8+ PVD synthesis parameters using statistical correlation, raising material reproducibility by ~15% and cutting failed experimental runs.
- Authored 7+ peer-reviewed publications and presented findings at 5+ national conferences; research cited in 3+ subsequent studies in the thermoelectric materials domain.

Tech stack: Python, Statistical Modeling, Data Pipelines, Causal Analysis

Research Intern | CSIR–National Physical Laboratory, New Delhi | India

Jan 2016 – Aug 2016

- Analyzed electrical and structural datasets from thin-film material experiments to evaluate material performance and reliability across 5 configurations.
- Contributed to building parameter-based predictive models for Resistive RAM behavior, identifying 4+ key variables influencing switching characteristics.
- Completed 100+ electrical and deposition measurements, calibrating sputtering parameters and generating structured datasets for material performance analysis.

PROJECTS

Dynamic Pricing Optimization using Machine Learning

- Analyzed 1,000 ride-sharing transactions across 3 demand scenarios (low, medium, peak) and found that just 2 features — rider-driver ratio and time-of-day — explained 80% of fare variance; this insight gave the operations team a simple, data-backed rule for when to apply surge pricing rather than relying on gut instinct.
- Tested 3 pricing models (Linear Regression, Random Forest, XGBoost) and identified XGBoost as optimal; the resulting model reduced pricing error by 24% over baseline, meaning fewer underpriced rides during peak demand and less overcharging during low demand — directly improving revenue yield and customer fairness simultaneously.

Twitter US Airline Sentiment Analysis using NLP

- Analyzed 14,640 tweets about 6 US airlines from February 2015 and found that 63% of all customer sentiment was negative — identified that United Airlines and US Airways together accounted for 46% of all complaints, with Customer Service Issues being the single largest driver of negative feedback at 20% of all negative tweets.
- Built and deployed a full-stack NLP application: trained 5 classification models (Logistic Regression, Naive Bayes, SVM, LSTM, BERT) with BERT achieving highest accuracy; built a live Streamlit web app with a real-time tweet sentiment classifier and interactive airline insights dashboard — deployed to Streamlit Cloud with a public URL, making findings immediately accessible to any stakeholder without technical setup.
- Translated findings into an actionable airline operations brief: fixing the top 3 negative reason categories could convert an estimated 40%+ of negative sentiment — directly quantifying the customer experience impact of operational improvements.

Predicting Ticket Prices for Big Mountain Resort

- Designed and assessed scikit-learn regression models to predict ski resort ticket prices based on operational features including vertical drop and night-skiing capacity.
- Applied EDA and SelectKBest feature selection to identify key predictors and reduce multicollinearity, cutting the feature set by 35% while maintaining predictive accuracy.

EDUCATION

Ph.D. in Physics (Material Science) Malaviya National Institute of Technology, Jaipur

Jun 2018 – Aug 2023

Masters of Science in Physics Punjabi University Patiala, Punjab

Jul 2014 – Jun 2016

Bachelors of Science Chaudhary Devilal University Sirsa

May 2011 – Jul 2014