

## Shrey Arora

(646)-637-1958 | <https://www.linkedin.com/in/shreyarora1> | [shreyarora@arizona.edu](mailto:shreyarora@arizona.edu) | <https://github.com/shreyarora150>

---

### EDUCATION

#### University Of Arizona - MS Statistics and Data Science

Aug 2022 – May 2024

Relevant courses: Statistical Machine Learning, Advanced Regression Analysis, Theory of Statistics, Neural Networks, SQL/No SQL Databases, Artificial Intelligence, Bayesian Statistics

GPA 3.56/4.0

#### Thapar University, Punjab - B.E Electrical Engineering

Aug 2015 – Jun 2019

Relevant courses: Object Oriented Programming, Numerical Analysis, Data structures and Algorithms, Optimization Techniques

GPA 7.62/10

---

### PROFESSIONAL EXPERIENCE

#### University of Arizona, AZ

Jan 2024 – May 2024

##### Graduate Teaching Assistant

- Graduate Teaching Assistant at department of Physics and Atmospheric Sciences. I taught PHYS 240 Electricity and Magnetism and took lab sessions for two sections.

#### Tesla, Fremont CA

Aug 2023 – Dec 2023

##### Data Engineering Intern

- Developed Lexicon, an **automated data dictionary mapping tool** integrated with Code-X CI/CD, utilizing **Apache Airflow** for deployment. **Achieved 85% accuracy** in mapping data-mart columns to pipeline sources and displaying results on ZIPLABS dashboard through **hierarchical recursive queries**.
- Worked on proof-of-concept to transition ETL data pipelines storing **Apache Kafka streaming data** from **Vertica** to **Clickhouse** database, optimizing resource usage and advancing toward a **scalable in-house solution**.
- Developed an **ETL pipeline** for packaging data mart table, alongside **ODS and STAGE layer** tables in Vertica database. Deployed the pipeline on **Apache Airflow** for efficient data processing and management.

#### American Express, India

Dec 2021 – Jun 2022

##### Analyst - Risk & Info Management

- Developed **forecasting machine learning models** using **Python(Numpy, Pandas, Sklearn)** for predicting transaction volumes and setting quarterly targets for the Sales and Account Management Team.
- Created new triggers, built data quality checks, analyzed **big data using Pyspark** and provided actionable insights for lost transaction volume, which led to a **30% reduction in annual lost charge volume**.
- Optimized **SQL queries and ETL scripts** that **reduced data pre-processing time by 2 hours**, which earned me a **Strategic Mindset Recognition**.

#### Prione (Amazon), India

Sep 2020 – Sep 2021

##### BI Analyst

- Conducted **root cause analysis** on success metrics of onboarded vendors, leveraging **exploratory data analysis** techniques and **A/B Testing** to understand the web behaviour of end customers which **improved annual Sales and Views by 40%**.
- Created **dashboards and visualizations**, and maintained **data pipelines** using **Python, SAS, Airflow and Tableau** that helped the pricing and performance insights team in data-driven decision-making in **tracking churns, sales and transactions**.
- Created **data visualizations** for weekly business reviews and other ad-hoc projects using **matplotlib, seaborn** and BI tools like **Tableau and QlikView**, which **enabled real-time visibility of metrics** for team leads and external stakeholders.

#### Happay (CRED), India

Jun 2019 – Sep 2020

##### Data Scientist

- Developed a **classification machine learning model** trained on transactions, subscriptions and user profiling features, performed **feature engineering** and **hypothesis testing**, added complexity to model as needed that **reduced churn by 28%**.
  - Created a **K-means clustering unsupervised ML model** that helped in prospecting clients for a new fuel card product by establishing priority scores which helped in **increasing sales conversion by 35%** from the expected number.
- 

### TECHNICAL SKILLS

**Languages** : Python (Numpy, Pandas, Scipy, Sklearn, Tensorflow), R, SAS, SQL, JavaScript, Matlab, HTML, VBA, C/C++

**Tools/Tech** : Machine Learning, Time Series, NLP, REST API, Agile, Git/Github, Hadoop, Airflow, Hive, DBMS, MongoDB

**Cloud** : Jenkins, Bitbucket, AWS Lambda, AWS API Gateway, EC2, Google Apps Script, dbt, Alteryx, Big Query

**Data Visualization** : Tableau, AWS Quicksight, Qlikview, Google Data Studio, Matplotlib, Seaborn, Power BI, MS Excel, Looker

**Databases** : MySQL, Postgresql, Redis, Vertica, MongoDB, ClickHouse, OpenSearch

---

### PROJECT EXPERIENCE

[Artificial Intelligent Pac-man](#) - Search Algorithms | Reinforcement Learning | Optimization | Python

- Created an AI enabled Pac-Man agent to optimize decision making in live game . Implemented agents using search algorithms like Depth-First Search, Breadth-First Search, A\* ,minimax, expectimax and reinforcement learning (Q-Learning) to make pacman learn through trial and error. Designed better evaluation functions to ensure better in-game performance.

[YouTube Comment Multi-Labeling](#) - NLP | Neural Networks | Text Multi Labelling | Hyperparameter tuning

- Trained a bidirectional neural network on Youtube comments data that inputs given text and predicts whether any of the given emotions are present: admiration, amusement, gratitude, love, pride, relief and remorse. Used hugging face's Roberta-Base tokeniser and pre-trained model to add embeddings in the model that achieved 84% macro f1 score on test-set.

[Huawei Advertisement CTR Prediction](#) - ML Modeling | Resampling | Ensembles | Boosting | Python | VS code | R Studio

- Modeled the click-through rates for mobile advertising. Large imbalanced dataset was analyzed, and resampling techniques were deployed with different classification methods to minimize class imbalance. The model was optimized and hypertuned to achieve an 82% f1 score with XGBoost.

[Database Design for E-Commerce Store](#)- MySQL | Snowflake | Normalization | Database Modelling

- Created conceptual, Logical and Physical models and designed a database in MySQL for an e-commerce store, used snowflake schema 2NF norm and junction tables to map many-to-many relationships. Leveraged advanced queries and stored procedures to create OLAP schema and tables and views.