

# Nishat Parveen

nishatconnect@gmail.com | 628.502.3712 | Mountain View, CA

[linkedin.com/in/nishat-parveen/](https://www.linkedin.com/in/nishat-parveen/) | [nishat-parveen.github.io](https://github.com/nishat-parveen) | [github.com/nishat-parveen](https://github.com/nishat-parveen)

## Professional Experience

---

### Pure Storage

Oct 2020-Present

*Data Analyst, Bay Area*

- Managing company expenditure on AWS resources based on key metrics and rightsizing; implementing cost cuttings by 30%.
- Deploying ETL on Airflow for AWS Cost analysis on Chartio (visualization dashboard).

### Checkpoint Software Technologies

Jan 2020-Jun 2020

*Data Science Intern, Bay Area*

- Applied machine learning techniques for cyber security of cloud serverless entities.
- Improved unsupervised anomaly detection models by clustering temporal data (ensemble of k-means) and using Isolation Forests resulting in false positive reduction from 25% to 0.07%.
- Synthesized weakly labeled data using SMOTE (Synthetic Minority Over-Sampling Technique) to validate model performance, achieving an F1 score of 80%.

### Honeywell Technology Solutions

*Tech Lead, Bangalore*

Jul 2016-Jun 2018

- Reduced the purchase cost of licenses by fitting ARIMA time series model in Python to predict license-based software usage for all engineering operations.
- Designed a chatbot using Long Short Term Memory Model (LSTM) for proprietary software installation and bug reporting.

*Senior Engineer, Hyderabad*

Jun 2018-Jun 2019

- Lowered the cost of handling the legacy billing system (20M USD revenue per year) for Satcom1, a startup acquired by Honeywell by creating an application using Django and SQL and integrating it with its billing portal.
- Decreased CPU usage by 70% and increased code maintainability by replacing heuristic-based flight crew alerting system with a Random Forest algorithm.
- Optimized cockpit information display by rewriting (C++) user interface to use only 3, rather than 4 monitors.

## Education

---

### University of San Francisco

Jul 2019-Jun 2020

*Master of Science in Data Science*

Courses: Linear Regression, Machine Learning, Deep Learning, Time Series Analysis, Reinforcement Learning, Distributed Computing, A/B Testing and Product Analytics.

### IIT Hyderabad

Aug 2012-May 2016

*B.Tech. in Electrical Engineering with Honors in Image Processing*

Courses: Linear Algebra, Probability, Machine Learning and Data Structures and Algorithms

## Projects

---

- Improved speaker recognition in noisy environment big data audio dataset by extracting features from .WAV files in Python and classifying using Random Forests in PySpark ML library with mean per class error of 0.196.
- Automated eye diagnosis of cataract patients (~5000 per day) with an image segmentation software using patch-based Particle Swarm Optimization in MATLAB. Paper published in *European Journal of Ophthalmology*.

## Technical Skills

---

Languages: Python, C++, SQL

Python Libraries: Scikit-Learn, NumPy, Pandas, TensorFlow, PyTorch, PySpark, Plotly, Seaborn

Technologies: AWS (EC2, EMR, EBS), Django, Git, Google Colab