

JEFFY MERIN JACOB

Permanent Resident of USA | Austin, TX 78736 | 815-575-7455

jeffyjac@usc.edu | <https://www.linkedin.com/in/jeffy-jacob/>

<http://www.scf.usc.edu/~jeffyjac/> | <https://github.com/jmerinj1996/My-Data-Science-Portfolio>

CAREER SUMMARY

Data Scientist • Machine Learning Specialist • Data Analyst • Software Engineer

- Graduate Data Scientist passionate about developing predictive models specifically, that help tackle societal problems. I am a data-driven storyteller characterized with a passion for coding and a zeal to take on challenges.
- Skilled in data cleansing, wrangling, visualization and machine learning.
- Worked extensively with data mining techniques to infer insights and make predictions in Python.

AREAS OF EXPERTISE

- *Machine Learning*- Python scikit-learn
- *Data Analysis*- Python pandas, NumPy, Beautiful Soup
- *Data Visualization*- Python seaborn, matplotlib, Tableau
- *Deep Learning*- TensorFlow, Keras, TFLearn
- *Database* - MySQL
- *Web Development* – Python Django, Flask, Dash
- *Programming languages* – Python, C++
- *MS Excel* – Pivot tables, Correlation
- *Statistics*– Hypothesis Testing, Regression
- *Cloud Services*- Google Cloud, AWS

PROFESSIONAL EXPERIENCE

USC Center for Artificial Intelligence in Society, Los Angeles, CA.

01/2019 – 08/2019

Student Intern

Part of a research group that worked on a preference elicitation algorithm to be used by the Los Angeles Homeless services to increase the chance of individuals exiting homelessness.

- Implemented Python Django web framework to display an interactive dashboard containing features of different policies.
- Performed data munging, exploratory data analysis and feature selection (RFECV) with cross-validation on data to determine best number of features accuracy wise when training a Logistic Regression model.

USC Information Technology Services, Los Angeles, CA.

05/2018 – 08/2018

Summer Research Intern

Part of the Data Networks and Operations team

- Designed a dashboard using Python Dash web framework and network measurement toolkit to visualize live packet moment and re-transmits between BWCTL servers thereby monitoring network performance over time.

Dowell Technologies, Chennai, India.

05/2017 – 07/2017

Intern

- Developed numerous simple to complex SQL queries involving procedures, functions, triggers for diverse business requirements.
 - Optimized queries using indexing strategies and altering database design.
-

EDUCATION

Masters, Computer Science (Data Science), Los Angeles, CA.

08/2017 – 12/2019

University of Southern California

Courses

Machine Learning - Data Mining - Artificial Intelligence - Database Systems - Geospatial Information Management - Information Retrieval and Web Search Engines - Web Technology - Analysis of Algorithms

Bachelors, Computer Science and Engineering, Chennai, India.

07/2013 – 05/2017

St. Joseph's College of Engineering, Anna University

PROJECTS

ASL Interpretation using Deep Learning - [see project](#)

07/2020 – 08/2020

- Designed a deep learning neural network in **TensorFlow** that classified images that belonged to 1 of 29 classes comprising the English alphabets A-Z in sign language. The network consists of 2 hidden layers and **SoftMax** to convert weights to probabilities. The optimizer used was **Stochastic Gradient Descent**.

Exploratory Data Analysis on Donor Choose dataset- [see project](#)

06/2020 – 06/2020

- An analysis on the kind of projects posted on DonorChoose.org to help donors find projects that inspire them the most. Interactive Pie charts and Histograms using **Plotly**.

Applications of Data Mining: Charity Matching – [see project](#)

03/2019 – 05/2019

- Built a **content based** and **collaborative filtering** recommendation systems to recommend relevant projects to donors in Oakland. Identified targeted donors who have previously donated to a cause and recommend similar projects to them.

Machine Learning for Data Informatics

01/2018 – 04/2018

- Improvement of Python code that created **expert variables** for a fraud detection algorithm. Initial code took 24 hours to run 100,000 records. The modified code used **NumPy** and **Pandas** and took 20 minutes to run 97,000 records in addition to 4 more expert features.
- Developed **decision trees**, **gaussian mixture models**, **SVMs** in Python as part of coursework with a purpose of developing a deeper understanding of the internal workings of various classifiers.

CERTIFICATIONS

- Statistics Foundations: 2 (07/2020 – 08/2020)
- Python for Data Science Essential Training (05/ 2020, LinkedIn Learning)
- Applied Machine Learning: Foundations (05/2020, LinkedIn Learning)
- Building a Recommendation System with Python Machine Learning and AI (05/2020, LinkedIn Learning)
- Data Science Foundations (09/2019, Lynda.com)
- Python Essential Training (12/2018, Lynda.com)
- Statistics Foundations: 1 (12/2018, Lynda.com)
- EMC Academic Associate (03/2016)