PAVLEEN KAUR

• +1(213)-326-4224 • San Jose, California • 19 Nov 1995
• pavleenk@usc.edu • github.com/pavleenkaur

EDUCATION

Master of Science in Data Science, University of Southern California, Los Angeles **May 2020**
B.Tech. Information Technology, Indira Gandhi Delhi Technical University, India **May 2018**

TECHNICAL SKILLS

Area of Experience : Data Analysis, Data Visualization, Business Intelligence, Machine Learning, Deep Learning, Data Mining, Database Management, Natural Language Processing, Web Scraping, Sentiment Analysis, Computer Vision, Software Development
Languages & Tools : Python, SQL, Tableau, D3.js, Trifacta, Weka, Apache Spark, Microsoft Excel, Microsoft Office, Oracle Cloud Analytics, C, C++, HTML, CSS, PHP, JavaScript, JQuery, JSON, Lua, Firebase, MongoDB, DynamoDB, Wordpress Content
Python Libraries : Scikit-learn, Tensorflow, Keras, PyTorch, NumPy, Pandas, Matplotlib, Plotly, OpenCV, NLTK, Plotly

EXPERIENCE

Genentech - Data Analyst Intern, *South San Francisco, California* **June 2020 - Present**

- Managing gRide (employee transportation program), and responsible for introducing route optimization and reservation forecasting
- Using Excel and Tableau to create preliminary cost saving routes for gRide and developing executive dashboards for the Site Services and Transportation Team to display trend reports
- Will be further deploying AWS and Google's Optimization Tools for the Vehicle Routing Problem, to fetch data, create feasible real time routes, monitor performance of fleet and report data driven results

NEC Laboratories America - Machine Learning Research Intern, *Princeton, New Jersey* **May 2019 - August 2019**

- Migrated NEC's Intrusion Detection and Classification System (IDS) from the in-house Lua based Machine Learning platform to PyTorch in order to improve access and understandability of the system
- Incorporated technical modifications to IDS, and decreased runtime by ~20%, while introducing functional programming for code generalization
- Refined and restructured incoming fibre optic cable data, attaining an improving F1 score by ~6%, and accuracy by ~12%, to 94.23%
- Introduced real time detection and classification in the system, with an output latency of 0.5 seconds and true positive rate of 0.98
- Incorporated the principle behind IDS for the prediction of coefficient of modal decomposition of a light beam, and published the novel findings in the leading conference SPIE Photonics West

Analytics Vidhya Pvt Ltd - Data Science Intern, *Delhi, India* **May 2017 - July 2017**

- Utilized Python, Tableau dashboards and Trifacta for supermarket sales analysis
- Reported results in 3 tutorial style articles on Tableau, that yielded over 50K views and increased user base by 2.7%
- Accelerated the monthly visitors by contributing towards researching machine learning algorithms and generating case studies and skill tests for the company's DataHack portal, used by 3M+ monthly visitors

PUBLICATIONS

- Pavleen Kaur, Giovanni Milione, Eric Cosatto, Philip Ji, NEC Labs. America, Inc., "Machine-learning orbital angular momentum spectra", SPIE Photonics West 2020, San Francisco, United States
- Pavleen Kaur, Payel Ganguly, Saumya Verma, "Bridging the Communication Gap : With Sign Language Conversion", 2018 IEEE/ ACIS 17th International Conference on Computer and Information Science (ICIS), 978-1-5386-5892-5

PROJECTS

Identifying Bias in Airbnb - *Python, Tableau, TensorFlow* **August 2019 - December 2019**

- Implemented a Convolution Neural Network to classify host images by race, with a balanced accuracy of 97%
- Analyzed sentiments on customer reviews, to identify racial and gender bias
- Used XGBoost to find feature significance and developed a fair price predictor that allocated prices to listings based on accommodation features

Art Recommendation - *Oracle Cloud Analytics, Python, Spark, Apriori, MLlib.* **January 2019 - April 2019**

- Developed a 2-way recommendation system to optimize artwork placement in the National Art Gallery of Singapore
- Exploited Apriori frequent item set generation to create pairs of visitors and the artworks they visited, further using Jaccard-based Locality Sensitive Hashing to find similarity between the various visitors to curate user based recommendations
- Visualized and generated patterns in the footprints of the visitors based on their past visits to the gallery, thereby curating item based recommendations of artworks to visit, with a standard deviation from the actual path trajectory of ~9%
- Presented top 5 recommendations for strategic location of the art to the gallery management team, with an RMSE (error) of 0.016

Sign Language Conversion (ASL) - *Python, OpenCV, Tensorflow* **August 2017 - May 2018**

- Interpreted sign language from images and videos using spatial and temporal learning, through the implementation of Recurrent and Convolutional Neural Network, with a mean accuracy of 92.3%
- Published the novel implementation and findings in the leading international conference ACIS 2018