Pavleen Kaur
+1(213) 326-4224
San Jose, California

pavleenk@usc.edu
github.com/pavleenkaur
https://www.linkedin.com/in/pavleen-kaur-usc-ghc2020/

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
Programming Skills : Python, C, C++, SQL, D3.js, Apache Spark, Scikit-learn, Tensorflow, Keras, PyTorch, NumPy, Pandas, Matplotlib, Plotly, OpenCV, NLTK, Plotly , HTML, CSS, PHP, JavaScript, JQuery, JSON, Lua, Firebase, MongoDB, DynamoDB
Tools : Tableau, Trifacta, Jupyter Notebooks, Weka, Github, Microsoft Excel, Microsoft Office, Oracle Cloud Analytics

EXPERIENCE

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

- Responsible for introducing efficiencies and cost reductions in gRide (employee transportation program) in response to COVID-19
- Spearheaded the development and execution of an optimum allocation algorithm on buses to routes under gRide; with ~45% increase in daily utilization rate, and reduction of ~31% in cost per seat and ~22% in overall costs
- Formalized a reservation forecasting system for gRide, based on moving averages, to predict weekly ridership with a mean absolute deviation of 1.105, and RMSE of 1.506

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

- Led the management and development of NEC's Signal Detection and Classification System (SDS)
- Migrated SDS from in-house Lua based Machine Learning platform to PyTorch to improve access and understandability of system
- Incorporated technical modifications to SDS, 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%, 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 principle behind SDS for prediction of coefficient of modal decomposition of light beam, and published novel findings in 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**

- Programmed a bias analysis, detection and resolution system using Airbnb Data
- 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**

- Designed a sign language identification and conversion algorithm for American Sign Language
- Interpreted sign languages 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