SHAURYA KHURANA

Vermillion, SD – 57069 | +1 (605) 202 9181 | shauryakhurana0@gmail.com

https://shauryakhurana.netlify.app/ https://www.linkedin.com/in/shaurya-khurana/ https://github.com/jamwine

EDUCATION

University of South Dakota **Vermillion, SD**

Master's in Computer Science, 3.6 *Jan 2020 – Jun 2021*

Relevant Coursework: Applied Mathematics for Machine Learning, Data Structures and Algorithms, Computer Vision, Pattern Recognition, NLP, Artificial Intelligence, Feature Engineering

WORK EXPERIENCE

University of South Dakota **Vermillion, SD**

Graduate Research Assistant *Jan 2020 – Jun 2021*

Tech Stack: **Python, Elasticsearch, Jira, GitHub, Streamlit, Lucidchart, Heroku, Bitbucket, Globus, Kibana, Travis CI, WorkFlowy, Confluence, Netlify**

- Designed a data mining pipeline for Biofilm Data & Information Discovery System, created REMAP to structure and sync data with Elasticsearch.
- Leveraged Smartsheets API to programmatically access hierarchical data and transformed it into user-interactive dashboard for researchers.
- Built a web application 2DMatchecker in Streamlit to compute 2D material characteristics using Logistic Regression and KNN Classifier.

Zenatix **Gurgaon, India**

Data Analyst *Feb 2019 - Dec 2019*

Tech Stack: **Python, Django, Docker, sMAP, scikit-learn, SciPy, statsmodels, Gitlab, NumPy, JSON, PostgreSQL, Plotly, Seaborn, Matplotlib, cufflinks**

- Exploratory data analysis - analyzed unstructured big-time-series from IoT sensors by data resampling, identified use cases by testing hypothesis with statistical tests, and benchmarked data by detecting anomalies using visualizations for scalability across customers.
- Developed a probabilistic statistical model to monitor the health of electrical assets, predict equipment breakdown, and schedule electrical equipment operations using Fourier Transforms, Signal Processing, and ARIMA with over 90% F-score.
- Clustered appliances by K-Means model to schedule time-based and temperature-based controls**,** saved energy and reduce costs up to 30%.
- Introduced a data extracting platform (DWYW) using ipywidgets for account managers, to automate the daily ad-hoc data requests.

Cognizant **Hyderabad, India**

Programmer Analyst *Nov 2016 - Feb 2019*

Tech Stack: **Python, AWS, RDS, SQLite, Java, Flask, MongoDB, Boto3, Tableau, Redash, ELK stack, Postman, Dynatrace, Puppet, Linux, YAML, Grafana**

- Integrated multiple sources of leases data into one system using Pandas and SQL, automated the process of loading DIF files into TRIRIGA SaaS.
- Accelerated Business intelligence reporting to assist clients with purchase-versus-leasing decisions by utilizing Python SciPy stack and Tableau.
- Innovated a data-driven model to determine the employee's incentive by integrating ticket data from ServiceNow using Google Data Studio.
- Upgraded the KLT framework to fetch customer surveys by designing REST APIs, and mapped feedback into the application by transitioning from a relational database to DynamoDB using the Agile approach.

Coursera **Remote**

Beta Tester & Mentor *Aug 2017 - Present*

- Acquired 80+ certifications from eminent universities and organizations in the field of technology, finance, and management by testing online courses before their launch. Providing feedback to course instructors to make sure the finished course is of high quality and free from errors.
- Mentoring learners worldwide with their issues through active participation in Coursera community and Discussion Forums.

SELECTED PROJECTS

Understanding Chest Radiographs abnormalities (Tuberculosis and Pneumonia) *Jan 2021*

- Utilized Image Processing techniques like segmentation, superimposing images, convex hull, canny edge detection and erosion/dilation to convert chest X-rays into image objects using scikit-image and OpenCV on Shenzhen Hospital X-ray dataset.
- Used Deep learning CNNs to train an image classifier with TensorFlow and Keras, investigated the performance - training time trade-off for architectures like VGGnet, Alexnet, Resnet with multiple loss functions, optimized algorithms, and hyperparameter tuning to achieve 85% AUC.

COVID -19 dashboard *Aug 2020*

- Scraped data from various publicly available sources on web, normalized it using Standard Scalers, and trained it to predict death and recovery rates worldwide with up to 90% accuracy using Polynomial Regression, GridSearchCV, and Cross-Validation scheme.
- Analyzed trends for countries using Heatmaps and Waffle Plots by utilizing visualization tools including Folium, pygal, Bokeh, and Streamlit.

Genetic Algorithms simulations *May 2020*

- Defined a schema to evaluate the fitness of randomly generated candidate solutions for parent solutions, crossover genetic information from parents to children, and then mutation of children to form the next generation of the population.
- Automated visual animations of GA process using Matplotlib to solve NP-complete problems - Travelling Salesman Problem and String Matching.