

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

MS in Computer Science, 3.6

Relevant Coursework: Applied Mathematics for Machine Learning, Data Structures and Algorithms

Vermillion, SD

Jan 2020 – Jun 2021

WORK EXPERIENCE

USD Biomedical Department, READS Tech

Software Engineer

Vermillion, SD

Jan 2020 – Jun 2021

Tech Stack: Python, Elasticsearch, Spark, Databricks, Airflow, Jira, Singer ETL, Globus, Kibana, Dask, Redshift, Streamlit, PyPika, S3, Heroku, GitHub

- 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 the user-interactive dashboard for researchers.
- Built a web application 2DMatchchecker in Streamlit to compute 2D material characteristics using Classification and Regression Tree (CART).

Zenatix

Data Analyst

Gurgaon, India

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

Programmer Analyst

Hyderabad, India

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 assisting 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 following the Agile approach.

Coursera

Beta Tester & Mentor

Remote

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.

PROJECTS/ TOOLS

Pre-processing Harmonized Landsat 8 and Sentinel-2 (HLS) Data (*gdal, NetCDF, ArcGIS, Hadoop, PySpark, GeoJSON, Vaex*)

Present

- Exporting and streaming HDF5 files into GeoTIFFs using Shapefiles, structuring metrics at defined frequency intervals by removing inconsistencies.

Artificial Intelligence to Save Cultural Heritage Documents (*OCR pyTesseract, Pillow, Computer Vision, Kraken*)

Mar 2021

- Document image analysis with pattern recognition and template-based matching to structure and extract information (texts and graphics).

Music Jambox (*BeautifulSoup, SVD, Random Forest, Regular Expressions, Feature Engineering, Web Scraping, Earsketch*)

Dec 2020

- Built a recommendation system to discover music, scraped artist's info, discography, song lyrics, detected Mp3 audio, and composed music.

COVID-19: Prediction and visualization (*scikit-learn, Folium, Dash, pyqal, Bokeh, Trend Analysis, Netlify*)

Aug 2020

- Extracted data from various available sources on web, normalized it using Standard Scalers, found best hyperparameters by tuning GridSearchCV.
- Predicted spread and recovery rates worldwide & country-wise with up to 90% AUC using Polynomial Regression and Cross-Validation scheme.

Genetic Algorithms simulations (*Fitness-score, Sampling, Crossover, Mutation, Matplotlib, FuncAnimation*)

May 2020

- Defined a schema to automate genetic process cycles visually for NP-complete problems like Travelling Salesman Problem and String Matching.

Edge Map Detection in Chest Radiographs (*OpenCV, TensorFlow, Keras, scikit-image, Convolutional Neural Networks*)

Feb 2020

- Utilized Image Processing techniques like segmentation, superimposing, convex hull, canny edge detection, and erosion/dilation for ROI in lungs.
- Investigated performance-training time and loss functions in Deep learning architectures AlexNet, ResNet, & VGGNet to train an image classifier.