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

WORK EXPERIENCE (6+ years)

Bombora   Reno, NV
Data Engineer III   *Jul 2021 - Present*

Tech: Python, GCP, Airflow, BigQuery, DataFlow, Dataproc, Kubernetes, Pub/Sub, Apache Beam, Jenkins, Apache Spark, MLflow, JFrog, Java, Datadog
- Directed the Conda to Pip migration project across the organization to support faster development, and increased overall efficiency by 95%.
- Orchestrated and ingested Intent data with Airflow periodically, bounded (batch) data using DataFlow (Apache Beam) and Pub/Sub, and unbounded (streaming) Events using Dataproc (Apache Spark and Map Reduce) into a data warehouse in BigQuery.
- Validated and refined data pipelines to ensure data quality issues with the Great Expectations tool.
- Originated the transfer of NetFactor/Sleuth system and IP-Identifier service into the Identity team to derive company data from the IP addresses.
- Automated the A/B comparison of new IP2C data sources and the Profile Resolution workflow of Merged data tables before scaling data.

USD Biomedical Department, READS Tech   Vermillion, SD
Senior Software Development Engineer   *Jun 2020 - Jun 2021*

Tech: Python, Elasticsearch, Databricks, PySpark, Apache Airflow, Singer ETL, Kafka, Globus, Kibana, Redshift, Streamlit, PyPika, Dask, Jira, Bitbucket
- 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 2DMatchecker in Streamlit to compute 2D material characteristics using Classification and Regression Tree (CART).
- Implemented an authentication gateway for the Globus platform using Python SDK.

Machine Learning Engineer/ Graduate Assistant   *Jan 2020 - Jun 2020*

Tech: Python, Docker, Bash, OpenCV, TensorFlow, Keras, scikit-image, Convolutional Neural Networks, SQL, Computer Vision, Pattern Recognition
- Developed highly scalable classifiers for Edge Map Detection in Chest Radiographs leveraging machine learning and rules-based models.
- Defined Region of Interest (ROI) in lungs by utilizing Image Processing techniques like segmentation, superimposing, convex hull, canny edge detection, and erosion/dilation. Investigated performance-training time and loss functions in Deep learning architectures to train image classifiers.

Zenatix   Gurgaon, India
Data Scientist   *Feb 2019 - Dec 2019*

Tech: Python, Django, Docker, scikit-learn, SciPy, statsmodels, Seaborn, Matplotlib, NumPy, Plotly, PostgreSQL, cufflinks, Gitlab, Sourcetree, sMAP
- Exploratory data analysis - analyzed unstructured big-time-series from IoT sensors by data resampling, identified use cases by testing hypotheses 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 by up to 30%.
- Introduced a data extracting platform (Do What You Want) using ipywidgets for account managers, to automate the daily ad-hoc data requests.

Cognizant   Hyderabad, India
Programmer Analyst   *Nov 2016 - Feb 2019*

Tech: Python, AWS, Flask, SQLite, Java, MongoDB, S3, Tableau, Redash, ELK stack, RDS, Boto3, Postman, Dynatrace, Puppet, Linux, YAML, Grafana
- Integrated multiple sources of financial data with Pandas and SQLAlchemy to automate 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 relational databases to a NoSQL database DynamoDB following the Agile approach.

Coursera   Remote
Beta Tester & Mentor   *Aug 2017 - Present*
- Testing online courses before their launch, and ensuring proper content including videos, audio, quizzes, and assignments.
- 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 and guiding them with career track.
- Actively participating in Discussion Forums for answering general queries and exploring new ideas.

EDUCATION

University of South Dakota   Vermillion, SD
MS in Computer Science, 3.7   *Jan 2020 - Jun 2021*
Relevant Coursework: Applied Mathematics for Machine Learning, Data Structures and Algorithms

Jaypee Institute of Information Technology   Noida, India
BTech in Electronics & Telecommunications, 3.74   *Jul 2012 - Jun 2016*

INDEPENDENT PROJECTS/ TOOLS

Automated Email Web Service (*Airflow, React, Redis, Docker, Jinja Templating, HTML, D3.js, Cookiecutters, Javascript)***   *Present*
- Creating the Email Newsletter Platform as a service for providing notifications service to subscribers and consumers.

Portfolio Construction and Analysis (*FastAPI, Airflow, Regression, Natural Language Processing, API Requests, Data Mining)***   *Present*
- Deriving Technical indicators from the time-series stock data, transforming scraped financial news into features using sentimental analysis in NLP, and ingesting fundamental indicators via APIs within daily Airflow jobs. Developing a platform for screening equities and maintaining API endpoints to fetch normalized big data.

.

Pre-processing Harmonized Landsat 8 and Sentinel-2 (HLS) Data *(gdal, NetCDF, ArcGIS, Hadoop, Spark, S3, GeoJSON, Vaex)*   *Apr 2021*
- Exported and streamed 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, Matplotlib, Computer Vision, Kraken)*   *Feb 2021*
- Document image analysis with pattern recognition and template-based matching to structure and extract information (texts and graphics).

Music Jambox *(BeautifulSoap, 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, pygal, Bokeh, Trend Analysis, Plotly, Netlify, Heroku)*   *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.

Designing Python, Data Science stack tutorials   Aug 2018 - Present
- Creating Jupyter Notebook tutorials for Python Data Analytical libraries, Statistics, and illustrating visualization examples.

ACHIEVEMENTS

- Acquired 100+ certifications from eminent universities and organizations in the field of technology, finance, and management.
- Certified the **Google IT Support Professional Specialization** to advance in networking, operating systems, troubleshooting, and security   Jun 2020
- Won the "Employee of the Quarter" award for developing the Do What You Want (DWYW) Platform, **Zenatix**   Sep 2019
- SPOT Award for contribution in achieving desired Business Results, **Cognizant**   Nov 2017

EXTRACURRICULAR & VOLUNTEER ACTIVITIES

Data Scientist, PEPROSYS   Jul 2022 - Present
CTO / Technology Innovator, DATAJOINS   Jun 2021 - Present
Business Consultant, VUOM ENTERPRISES   Jan 2021 - Present
Teaching Assistant, UNIVERSITY OF SOUTH DAKOTA   Aug 2020 - Dec 2020
Summer Research Intern, ALCATEL LUCENT   Jun 2015 - Jul 2015
Research Intern, CENTRAL ELECTRONICS LIMITED   Jul 2014 - Aug 2014
Digital Marketing Intern, GRADESTACK   Feb 2014 - Mar 2014
Business Analyst Intern, HAPPAY   Jul 2013 - Oct 2013