

Shailesh Kumar

kumar.shailesh1597@gmail.com | +917007786069 | LinkedIn: [shailesh-kumar-2171a814b](#) Website: [shaileshkumar97](#)

EDUCATION

Maulana Azad National Institute of Technology

Master of Technology in Artificial Intelligence; CGPA: 7.95/10

Bhopal, India

2019 - 2021

Shri Ramswaroop Memorial College of Engineering & Management

Bachelor of Technology in Electrical Engineering; Percentage: 74.14 %

Lucknow, India

2015 - 2019

- Gold Medalist, Best Project, Electrical Engineering Department, 2019.

TECHNICAL SKILLS

- **Languages:** Python, SQL, HTML/CSS
- **Frameworks:** FastAPI, Streamlit, Dash, Flask
- **Developer Tools:** Docker, Git, Visual Studio
- **Databases:** PostgreSQL, Redis, MySQL, AWS S3, Snowflake, Azure Blob, MongoDB, SQLite
- **Big Data:** Spark, Delta Lake, Data Lake (Hudi, Hive)
- **Libraries:** Pyspark, PyTorch, TensorFlow, Keras, Pandas, NumPy, Matplotlib, Seaborn, Scikit-Learn, OpenCV, PIL, Librosa, NLTK

EXPERIENCE

Data Scientist

katonic.ai (June 2021 - present)

- **Client Project-1:** Cloud Agnostic Model Access to non-AWS customers.
 - **Problems/Concerns:** Don't have access to the model artifacts built in sagemaker which can be exported for deployment into other cloud envs.
 - **Result:** Deployment and Maintenance to enable Cloud Agnostic Model Access.
 - **Remarks:** Utilized **253** Pre-built models for non-AWS customers with Time for Real-Time Inference of **under 4 mins** & Batch Transform Inference of **under 16 mins**.
- **Client Project-2:** Automate the Video Labeling Model (end-to-end Pipeline) Clustering (content & users) and Time to Process all.
 - **Problems/Concerns:** Time to process all **47000+ videos** (Data Source: AWS S3) takes **4+ hrs**.
 - **Result:** End-to-end Kubeflow Pipeline integrated with incoming data source, Time to process all **47000+ videos** reduced to **57 mins**.
 - **Remarks:** Identified performance issues bugs and suggested Optimization steps to reduce processing time.
- **Project-3:** Created a following modern, enterprise-ready **MLOps Python SDK**:
 - **Feature Store:** To manage end-to-end life-cycle of features & integrate with existing data stores, feature pipelines, and ML platforms.
 - **Connectors:** To access the data from different databases/warehouses and stores to a given destination.
 - **FileManager:** To access, store and update/manipulate objects within the file browser.
 - **Pipeline:** To convert an existing notebook into a Kubeflow pipeline.
 - **AutoML and Data Drift Detection:** Reviewed and Updated the complete packages for MLOps.
 - **Model Explainability:** To detect, monitor the drifts quality in the Batch and Production data.
 - **Log:** To quick register the trained models with mlflow in platform for deployment to the production environment.
- **Monitoring & Deployment:**
 - Collaborated with the Product Team on Model Monitoring to **Review and Update the code for key metrics & better performance**.
 - Implemented and Documented Swagger UI for **Model Monitoring APIs Handling**.
 - Partnered with Platform Team to Improve & Manage **Authorization & Data Handling on Model & App Deployment in Production**.
- **APIs & UI:**
 - Built **Backend APIs** in FastAPI & UI(beta) in HTML/CSS for **Feature Discovery/Search in Feature Store**.
- **Web Apps Development:**
 - Built, Updated, Documented and Deployed **40+** Web Application on different domains & use cases for Platform Accelerator.
 - Lead Data Science Team on **135+** Web Application Development on different domains & use cases.
- **Documentation:**
 - Created and Updated complete Python SDK Documentation (Internal & External).
 - Implemented and Documented Example Notebooks Tutorials for all sub-packages for Python SDK.
 - Created, Updated and Lead Team for SDK, User Guide and Getting-Started Documentation for Platform.
 - Designed and Implemented Architecture/Workflow Diagram for Feature Store (Backend + Frontend), Model Explainability (Backend + Frontend), Model Monitoring, Model Deployment and App Deployment.

OPEN SOURCE PROJECTS

- **Project:** [ExplainIt](#), a modern enterprise-ready business intelligence web application.
- **Features:**
 - **Analyze Drift** in the existing data stack (Features & Targets).
 - Prepare a very Short Summary of training & production data.
 - Perform **Quality Checks** on the data.
 - **Analyze in-depth relationship** between features & target.

INSTRUCTOR/WORKSHOPS

- Instructor for **Katonic MLOps Certification Course** on Udemy with **4.3** rating & **600+** enrollment.

INTERNSHIPS

Computer Vision & IoT Engineer

The Sparks Foundation (Mar 2021 - Apr 2021)

- Implemented a **MobileNet & SSD Algorithms with OpenCV's DNN** module to Detect Multiple Objects and Persons in the Image.

Data Analytics Consultant

KPMG (Sep 2020 - Oct 2020)

- Implemented data quality assessment and completeness in preparation for analysis.
- Implemented high value customers targeting based on customer demographics and attributes.
- Used visualisations to present insights.

PROJECTS

- **Blindness Detection:** Implemented some of the latest deep learning algorithms to help doctors diagnose various diseases.
 - Trained **ResNet101 & ResNet152** on **Blindness Retinopathy dataset** to classify Blindness Risk to prevent lifelong diseases. These models may be used to detect other sorts of diseases in the future, like glaucoma and macular degeneration.
- **Speech Enhancement:** Built a Speech Enhancement module to **Minimize Environmental Noise**.
 - Trained **U-Net**, a Deep Convolutional Autoencoder with symmetric skip connections used to denoise spectrograms.
 - Got a training loss of **0.002129** and a validation loss of **0.002406** on **5000** samples of clean voice and **1900** samples of noise.
- **StackOverflow Survey:** Data Analysis to explore the overall programming community. The **dataset contains 64k** responses to an annual survey conducted by StackOverflow.

ACHIEVEMENTS

- Won 2nd prize in Robo-Soccer competition in Gantavya'18, technical festival of SRMCEM Lucknow.
- Won 3rd prize in HurdleMania competition in Technex'18, technical festival of IIT BHU.
- Won 3rd prize in Robo-Race competition in Tech-Carnival'18, technical festival of AIMT Lucknow.

CERTIFICATIONS/COURSES

- AI Product Manager | Udacity (Nov 2021)
- Artificial Intelligence, Machine Learning & Deep Learning Program | Letsupgrade (May 2021)
- Data Structures and Algorithms in Python | Jovian.ai (Apr. 2021)
- Deep Learning with PyTorch: Zero to GANs | Jovian.ai (Jan. 2021)
- Data Analysis with Python: Zero to Pandas | Jovian.ai (Oct. 2020)
- Machine Learning | Stanford & Coursera (Oct. 2020)
- Introduction to TensorFlow Lite | Udacity (May. 2020)
- Introduction to TensorFlow for Deep Learning | Udacity (Apr. 2020)

POSITION OF RESPONSIBILITY

Gantavya 2016 - Technical and Entrepreneurial Festival

Aug. 2016 - Mar. 2017

Grobots Club

SRMCEM Lucknow

- Successfully organized Robo-Soccer, Robo-Race Event together for the first time in Gantavya.
- Organized workshops and Seminars for technical events.

Abhivyakti 2016 - Cultural Festival

Aug.2015 - Mar. 2016

Security Cell

SRMCEM Lucknow

- Successfully organized Xero, Starnight together and coordinated with different clubs throughout the festival in Abhivyakti.
- Promoted security awareness through seminars.

INTERESTS

- Rubik's Cube Solving
- 3D Animation
- Open Source Projects in AI, Machine Learning, Computer Vision, MLOps