

# Harshit Rai

---

[harshitrai68@gmail.com](mailto:harshitrai68@gmail.com)

| +918452904404

| [LinkedIn](#)

| [GitHub](#)

## EXPERIENCE

### DATA ENGINEER | SHELL INDIA | JULY 2019 - PRESENT

- Developed and deployed **end-to-end data pipelines** for three products on Microsoft Azure and Cloudera (before migrating to MS Azure) as a part of **The Geneva Initiative** (under Downstream Trading & Supply).
- Hands-on experience in building **ETL pipelines** using Azure Data Factory, Databricks, Data Lake Storage and Synapse Analytics.
- Collaborated with business analysts, trade analysts, data and solution architects to **formulate and implement business logic** across data pipelines.
- Led a **Continuous Improvement** initiative within the product team to optimize existing data pipelines and development processes.
- Designed and developed **Data Quality Control** mechanism to ensure data integrity and reduce business downtime.
- Created **Alteryx** workflows to automate reference data creation for data pipelines.
- Led the Machine Learning (NLP) team in Project EVA which adds a value of \$400,000/year to the organization.
- Trained cutting-edge **transformer-based NLP models** like BERT, DistilBERT, XLNet and T5 using **HuggingFace** library on **Azure ML Studio**.
- Built **ML pipelines** and deployed the trained models on Azure Container Instance (ACI) and Azure Kubernetes Service (AKS).
- Leading an NLP Proof of Concept on **AWS SageMaker** to onboard it as a preferred ML platform.
- Developing a user-interface for project GreenNLP using **ReactJS** and deployed it on **Azure App Service**.
- Leading an **organization-wide Python Upskilling Series**.

### FOUNDER | QUSTAC TECHNOLOGIES | JAN 2018 – JUNE 2019

- Led **business and software development teams** of 15+ employees and interns.
- Designed and implemented **software architecture for two AI products, dtoxd.ai and Balanced**, dealing with issues like deteriorating mental health, and exposure to cyber bullying and explicit content.
- Used **transfer learning** to train **Computer Vision** (MobileNet V1 & V2, EfficientNet, Xception and NasNet) and **Natural Language Processing** (RNN, LSTM and GRU) models using Keras and Tensorflow.
- Deployed these trained models on Google Cloud Platform (GCP) and Heroku.
- Led research teams on topics like Object Detection, Image Classification, Text Classification and Model Performance.
- Performed user data analysis using Python and data visualization libraries for deriving insights to drive business decisions.
- Developed and deployed web-applications using ReactJS on Google Cloud.
- Developed desktop applications using ElectronJS, C# and Python.
- Mentored 12 interns in their engineering final year project research.

## EDUCATION

Institution Name	Course	Score
Fr. C. Rodrigues Institute of Technology, Navi Mumbai (University of Mumbai)	B.E in Information Technology	8.51 CGPA
Royal Junior College of Science & Commerce, Dombivli (Maharashtra State Board)	HSC	79%
St. Therese Convent High School, Dombivli (Maharashtra State Board)	SSC	90.5%

## SKILLS

- Data Engineering, Machine Learning, Natural Language Processing, Computer Vision, MLOps, ETL, CI/CD.
- **Programming Languages**
  - Regular: Python, JavaScript
  - Past Experience: PHP, C, HTML (scripting language) and CSS (style language)
- **Tools & Technologies:** Databricks, Apache Spark, Git, SQL Server, TensorFlow, Alteryx, Azure Data Factory, Data Lake Storage and Synapse Analytics, HuggingFace, React
- **Interests:** React, Apache Kafka, C#

## CERTIFICATIONS

- Machine Learning Engineer using Microsoft Azure Nanodegree by Udacity
- Tensorflow in Practice by Deeplearning.ai
- Data Engineering with Google Cloud Specialization by Coursera
- Azure Data and AI Fundamentals (AZ-900, AI-900 and DP-900)

## PERSONAL PROJECTS

- **Explicit Content Censor**
  - A solution using Computer Vision and NLP to prevent cyberbullying and undesired exposure to explicit media.
- **Faculty Database Management System**
  - A college project aimed to eliminate the paperwork involved in maintaining the faculty records.
- **Agnel Online**
  - A college project aimed to eliminate the paperwork involved in maintaining the faculty records.

## ACHIEVEMENTS

- Won the CitiusTech #ChangeIndiaChallenge Hackathon (2019)
- Won 1st prize at National Level Project Competitions at Terna CoE and RAIT, Mumbai (2019).
- Won 1st prize in IRIS 2015 at Royal Junior College (Featured in newspapers and live television interviews for developing a tech to charge your phone while walking.)