# Vaibhav Gulati

⌂ Haryana, India  ☎ +91-9717295216  ✉ gulvaibhav20@gmail.com  ◯ GitHub  in LinkedIn

## Skills

| | |
|---|---|
| **Core:** | Data Engineering, ETL, Data Modelling, Large Language Models (LLMs), Generative-AI, System Design Natural Language Processing (NLP), Backend Development, RESTful APIs, Data Structures & Algorithms Object Oriented Programming, Design Patterns, Technical Content Writing, Research |
| **Languages:** | Python, Java, JavaScript, C/C++ |
| **Frameworks:** | PySpark, OpenAI, RASA, Numpy, Pandas, Plotly, Sktime, SpringBoot, Airflow, FastAPI, Django |
| **Databases:** | MySQL, SparkSQL, PostgreSQL, OracleDB, MongoDB |
| **Tools:** | Microsoft Azure, Git/GitHub, SonarQube, Postman, Redash, Docker, VS-Code, Jupyter, JIRA |

## Professional Experience

**Incedo Inc.**                                                                                                 **Jul 2022 − Present**
*Senior Software Engineer - Incedo Lighthouse™ (Data Focused)*                                   *Gurugram, India*

- Led the development of a **RASA**-based conversational AI chatbot that enhanced business operations visibility, enabling users to create KPI-Trees using **LLMs** and driving **40%** efficiency improvement in decision-making processes.
- Developed and optimized batch processing workflows with **Python, PySpark**, & **SparkSQL** for decision automation algorithms (Outlier detection, Cohort Analysis etc), achieving **5x** performance improvement over legacy systems.
- Spearheaded the database migration from **PostgreSQL** to **OracleDB**, supporting multi-database support via **Spring Hibernate ORM**, and ensured **99.9% data accuracy** with automated data sanity checks during migration.
- Designed and maintained backend infrastructure (**Java, SpringBoot, RESTful APIs**) for an internal request management system, improving data visibility and user engagement by **25%** through streamlined processes.
- Automated **JIRA** worklog monitoring by developing a Python-based reporting tool, generating real-time capitalization reports and reducing manual tracking time for team managers by **5+ hours weekly**.
- Designed and implemented a **SonarQube** code quality dashboard using **Redash**, **Python**, and **Airflow**, incorporating policy-as-code principles to achieve **80%** code coverage and enhance platform reliability.
- Developed dummy dataset generation scripts to simulate real-world data for testing, ensuring platform stability and accurate validation of decision automation algorithms for various business use cases, reducing testing time by **20%**.
- Authored comprehensive user walkthroughs, technical, and integration documentation each sprint release, improving deployment efficiency and saving over **15 hours per integration cycle** for development teams.

**Skills:** Python, PySpark, Gen-AI, RASA, Java, SpringBoot, Airflow, PostgreSQL, OracleDB, Redash, SonarQube

**Scaler | (10 Months)**                                                                                   **Nov-2021 − Aug-2022**
*Freelance Technical Content Writer*

- Demystified complex topics (**OOPS, design patterns**) through **27** engaging articles, empowering **20K+** learners on the Scaler Topics platform. [**Link**]

**Skills:** Technical Content Writing, OOPS, Design Patterns, SQL, Java, Data Structures & Algorithms

**Warpspeed Technologies Pvt. Ltd. | (2 Months)**                                             **Aug-2021 − Oct-2021**
*Backend Developer Intern*

- Spearheaded the development of **Queless**, a community-driven online ordering platform, utilizing **Django, RESTful APIs**, and **MySQL**.

**Skills:** Python, Django, Backend Development, RESTful APIs, Postman

## Research Projects

**Automated Soft Skills Evaluator**                                    <u>**Publication Link**</u>
*Published On: 01 Sept 2023*                    *Project Development Duration: 1 year (2021-2022)*

- Developed an automated system to assess the soft skills of an individual (using recorded webcam data). By analyzing frontal posture and vocal characteristics, the system estimates non-verbal and verbal confidence scores. Key features:
  * *Posture Analysis:* Utilized the <u>**MoveNet-Thunder**</u> skeleton estimation model to assess non-verbal confidence of an individual by measuring visible frontal posture angles such as shoulder alignment, neck tilt, and arm abduction.
  * *Disfluency detection:* Assessed verbal communication by identifying pauses, filler words, and word repetitions using a combination of **noise-reduction algorithms**, **Google Speech-to-Text model**, & **NLP** techniques.
  * *Combined Soft Skills Evaluation:* Integrated posture and speech analysis scores to generate a soft skills evaluation, validated against thresholds derived from the **Librispeech** dataset and ergonomic posture studies.
- **Tech:** Python, Posture Estimation, Speech-to-text, NLP, Computer Vision, Noise Reduction, Tensorflow

**Extractive Text Summarizer**                        <u>**Publication Link**</u> | <u>**GitHub Link**</u>
*Published On: 11 Jan 2023*                    *Project Development Duration: 8 months (2020-2021)*

- Developed a hybrid algorithm for extractive text summarization process that integrates the strengths of lexical-based **TextRank** algorithm and bag-of-words-based **BM25+** algorithm for improved summarizing accuracy.
- Employed a normalized similarity matrix to enhance the quality of the generated summaries and performed a comprehensive evaluation using standard **ROUGE metrics** to validate the effectiveness of the proposed method.
- **Tech:** Python, NLP, Textual Analysis, Machine Learning, TextRank, BM25+, NLTK, ROUGE

## Other Projects

**KOL Data Standardization (KOL Profiling)**                                    <u>**GitHub Link**</u>
*Field: Data Engineering (ETL)*                    *Project Development Duration: 1 month (2022)*

- Developed an **ETL** pipeline that extracted data from various sources, transformed it using business rules and master data, and loaded it into a standardized format. This enabled efficient data management and analysis, supporting healthcare & pharmaceutical companies in identifying & engaging with KOLs (Key Opinion Leaders) more effectively.
- Utilized **Microsoft Azure Data Factory**, **Azure Data Lake Storage**, & **Azure Functions** to implement cloud-based KOL Profiling solution to ensure event-based batch-processing & standardization of large KOL datasets.
- **Tech:** ETL (Extract, Transform, Load), Microsoft Azure Data Factory, Azure Functions, Data Modeling, SQL

## Education

**Bharati Vidyapeeth's College of Engineering**                                    **2018 - 2022**
*Bachelor of Technology - Computer Science & Engineering (CGPA: 9.364)*                    *Delhi, India*

**Kendriya Vidyalaya A.F.S.**                                    **2017 - 2018**
*Senior Secondary - 12th Standard (Percentage: 91.1%)*                    *Tughlakabaad, Delhi, India*

**Kendriya Vidyalaya No.1**                                    **2015 - 2016**
*Secondary - 10th Standard (CGPA: 10.0)*                    *Faridabad, Haryana, India*

## Achievements

- Incedo Innovation Customer Excellence award 2023
- Google IT Automation - Professional Certification   | **<u>Credly Link</u>**
- Microsoft Certified: Azure Fundamentals (AZ - 900) | **<u>Credly Link</u>**