

# Shivansh Verma

Bengaluru, Karnataka | P: 8800573508 | [shiv.v10899@gmail.com](mailto:shiv.v10899@gmail.com) | LinkedIn: <https://www.linkedin.com/in/shivansh-verma05/> | Github: <https://github.com/shivanshv9911>  
Portfolio: <https://shivanshv9911.github.io/shivansh-verma-portfolio/>

## PROFESSIONAL SUMMARY

---

Bioinformatics Data Analyst with 2+ years of experience building scalable NGS analysis pipelines and dashboards to support experimental biology teams. Proficient in Python, Snakemake, DESeq2, Google Cloud, and data visualization tools. Known for automating time-intensive workflows and integrating qPCR, RNA-seq, HybELISA, and Caspase-Glo assay data to generate actionable biological insights. Adept at collaborating with global teams to accelerate research timelines in drug discovery.

## EDUCATION

---

**Indraprastha Institute of Information Technology Delhi (IIITD)**

New Delhi,

**M.Tech – Computational Biology**

Aug 2021-Aug 2023

CGPA : 9.0/10

**Thesis Title** : Analysis of Gene Expression and Structural Variations in Common Genes Across Populations: Implications for Cardiovascular Disease and Reverse Cholesterol Pathway

**Noida Institute of Engineering and Technology (NIET)**

Greater Noida, UP

**B.Tech – Biotechnology**

Jul 2017 - Jul 2021

CGPA : 8.9/10

## PROFESSIONAL WORK EXPERIENCE

---

**Creyon Bio, India (Startup)**

Bengaluru, Karnataka

Data Analyst / Research Associate – Data Analysis

Aug 2023 – Present

- Designed and implemented automated workflows for internal assays (qPCR, HybELISA, RNA-seq, Nanopore) using **Python, Snakemake, and Streamlit**, accelerating analytics and decision-making.
- Built modular, end-to-end pipelines for DGE analysis, Nanopore data Analysis including QC, and assay processing, integrated into a Snakemake framework with cloud storage on Google Cloud.
- Developed a **Streamlit dashboard for qPCR and HybELISA assay**, reducing manual reporting time by 10 hours per week.
- Cleaned, categorized, and stored raw datasets, removing outliers and **ensuring high-quality analysis** inputs for ML models.
- Consolidated unstructured **in vivo data from 20+ sources and cleaned**, categorized, and stored raw datasets, which established structured formats with embedded QC metrics, accelerating development and machine learning deployment.
- Documented dashboard/report specifications, contributing to system scalability and reproducibility.
- Collaborated with cross-functional R&D teams (India/US) on experimental design, data integration strategies, and platform development.
- Worked closely with biologists and data scientists to interpret experimental results and iterate on assay design in real time.

## PROJECT WORK

---

- Comprehensive Analysis of RNA Sequencing, Connectivity Mapping, and Molecular Docking for Cardiovascular Disease Therapeutics**

**Guide:** Dr. Arjun Ray (Associate Professor, IIITD)

- Analyzed RNA-seq data to identify differentially expressed genes and pathways.
- Applied Connectivity Mapping to discover small molecules capable of reversing disease signatures.
- Validated drug candidates through molecular docking simulations.

- **Tools:** AutoDock, AutoDock Vina, DESeq2
- **Key Metric:** Identified top 10 candidate molecules with strong binding affinities.

## 2. Automated Web Scraping for Protein Dataset Development for Machine Learning Models

**Guide:** Dr. Arjun Ray (Associate Professor, IIITD)

- Built a BeautifulSoup-based web scraper to extract structured protein metadata from the RCSB database.
- Created a curated dataset ready for machine learning model training and evaluation.
- **Tools:** Python (BeautifulSoup)
- **Key Metric:** Extracted and cleaned 5,000+ protein entries.

## 3. Differential Gene Expression and Pathway Enrichment Analysis in the Reverse Cholesterol Pathway

**Guide:** Dr. Arjun Ray (Associate Professor, IIITD)

- Developed Bash scripts and Python workflows for bulk RNA-seq preprocessing and differential expression analysis.
- Conducted gene set enrichment analysis focusing on cholesterol metabolism pathways.
- **Tools:** Bash, Python (Seaborn, Matplotlib), DESeq2
- **Key Metric:** Identified 200+ differentially expressed genes associated with cholesterol metabolism.

## 4. Classification of Protein Interaction Patterns Using Machine Learning

**Guide:** Prof. GPS Raghava (Head & Professor, IIITD)

- Developed machine learning models to classify interacting vs. non-interacting protein patterns.
- Deployed the trained model into an interactive web application using Streamlit.
- **Tools:** Python (Scikit-learn, Streamlit)
- **Key Metric:** Achieved 85% classification accuracy on test datasets.

## ADDITIONAL

---

### Technical Skills:

- **Programming & Data Analysis:** Python, R, SQL, Pandas, NumPy, Jupyter Notebooks
- **Bioinformatics Tools & Pipelines:** Snakemake, Nextflow, DESeq2 (Python), FastQC, MultiQC, NanoPlot, STAR, HISAT2, featureCounts
- **Cloud & Infrastructure:** Google Cloud Platform (BigQuery), Docker, Git, Linux, Bash
- **Data Visualization & Dashboarding:** Looker, Streamlit, Seaborn, Matplotlib, Excel (Advanced)

**Awards:** 3rd Prize – Startup Conclave (IMS Ghaziabad, 2018); Department Topper Award (2020); TCS NQT Qualified (2020); GATE-2021 (Biotech).

**Languages:** Fluent in English and Hindi; Beginner in German

**Certifications & Training:** GCP Data Engineering (Coursera), Docker & Kubernetes (Udemy), Data Science with Python (IBM)

**CONFERENCES:** Attended illumina Genomic Summit, Bengaluru (30th August 2024)

## ACTIVITIES

---

### TEACHING ASSISTANT

Delhi

- Worked as Teaching Assistant in IIITD.

### STUDENT WELFARE

#### Committee Member

Greater Noida, UP

- Editorial member of Department of Biotechnology(NIET).
- Dean student welfare Coordinator of the Department of Biotechnology(NIET).
- Organized and advertised 5+ yearly departmental events with 50+ participants from all years in the department.
- Managed the launch of a new departmental magazine to improve flow of information and organization of events.