

**LEENA VIG**

DELHI 110070

EMAIL: LEEVIG1234@GMAIL.COM

MOBILE: +91-9818186338

LINKEDIN [HTTPS://WWW.LINKEDIN.COM/IN/LEENA-VIG-4BB57839/](https://www.linkedin.com/in/leena-vig-4bb57839/)

**Objective:** Highly motivated and analytical Data Scientist. Seeking to leverage my expertise in data analysis, machine learning, and statistical modelling to contribute to innovative projects and drive data-driven decision-making in a challenging industry role. Aim to integrate my knowledge and experience of teaching undergraduate students with the acquired skill set of Data Analysis and programming to assist in bringing forth solutions to burning problems faced by different sectors in the industry.

**EDUCATION:**

- Ph.D. in Molecular Biology from DR. B. R. Ambedkar Center for Biomedical Science, University of Delhi, Delhi in 2012.
- Masters of Science in Biology from University of Iowa, Iowa, USA, 2002.
- Masters of Science in Biotechnology from IIT, Mumbai in 1997.
- Bachelors of Science in Chemistry (Honours) from St. Stephen's College, University of Delhi, 1994.
- Successfully completed online Master's Program 'Artificial Intelligence Engineer' (offered by Simplilearn) having completed all the mandated course requirements and industry projects with distinction on January 23, 2022.
- Successfully completed online Master's Program 'Data Scientist' (offered by Simplilearn) having completed all the mandated course requirements and industry projects with distinction on August 22, 2022.

**EXPERIENCE:**

- Data Associate at Empire Analytics, Gurgaon, India (November 2023 till date)
- 10 years of teaching to Biochemistry(Honours) undergraduate students at University of Delhi (2012-2022). Subjects taught : Molecular Biology, Genetic Engineering, Bioenergetics, Metabolism. Organized many student training workshops, seminars

and extracurricular student activities as part of teaching responsibilities along with active participation in institute administrative roles.

- Served as Junior Scientist at Nicholas Piramal Pvt. Ltd. at Institute of Genomics and Integrative Biology, Delhi. My responsibilities as a Junior Scientist at IGIB included management and processing of blood samples from diabetic patients collected from various hospitals in Delhi and Mumbai with the aim to unearth novel single nucleotide polymorphisms in the candidate genes of diabetic patients in the Indian population. (May 2003 to Aug. 2004)

## SKILLS:

Data Analysis: Proficient in cleaning, transforming, and analyzing large datasets using Python.

Machine Learning: Skilled in developing and implementing machine learning algorithms for predictive modelling and pattern recognition.

Statistical Modelling: Experienced in applying statistical methods and hypothesis testing to derive insights from data.

Data Visualization: Proficient in creating informative and visually appealing data visualizations using tools such as **Tableau and Matplotlib**.

Natural Language Processing (NLP): Knowledgeable in NLP techniques for sentiment analysis and text classification.

Deep Learning: Familiar with neural networks, convolutional neural networks (CNN), and recurrent neural networks (RNN) for image recognition and sequence modelling.

Programming: Proficient in Python for data analysis and model development.

Big Data Technologies: Familiarity with Hadoop, Spark, and distributed computing for handling big data.

## EXAMPLES OF DATA SCIENCE PROJECTS COMPLETED:

**Project 1: Healthcare data set analysis project:** Aim of the project was to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements from an original data set obtained from the National Institute of Diabetes and Digestive and

Kidney Diseases, USA. Several constraints were placed on the selection of these instances from a larger database. The data set consisted of several medical predictor variables (BMI, insulin level, age, and so on) and one target variable. Analysis included descriptive analysis of the provided variables, checking whether data is balanced or not, creating scatter charts between the pair of variables to understand the relationships, performing correlation analysis and visually exploring the data using a heat map. Appropriate classification algorithms were applied to build a model followed by comparison of various models. Classification reports were generated for various models by analyzing sensitivity, specificity, AUC (ROC curve) and other relevant metrics. Finally, a tableau dashboard was created using appropriate chart types and metrics as useful for the business. ([https://github.com/LeenaVig/Healthcaare\\_data\\_analysis](https://github.com/LeenaVig/Healthcaare_data_analysis))

**Project 2: Financial data set analysis project:** Aim of the project was to identify fraudulent credit card transactions in a credit card company so that the customers are not charged for items that they do not purchase. Specific tasks completed in this project included exploration of the financial data of the company, tackling of the class imbalance problem presented by the data set by using techniques of oversampling and under sampling. Various models like Naive Bayes, Logistic Regression, SVM, Random Forest and XGBoost were used to model the data and appropriate metrics were used to evaluate the model performance. Artificial Neuronal networks were also used to model the data and their performance was compared with traditional ML models. ([https://github.com/LeenaVig/Financial\\_fraud\\_detection?search=1](https://github.com/LeenaVig/Financial_fraud_detection?search=1))

**Project 3: Analysis of factors effecting University student admission:** Aim of the project was to analyse the historical data given and determine the key drivers for admission to University in USA. The specific tasks completed in this project included finding missing values and outliers in the given data set, correcting the data types in the data set provided, normalization of the data, using variable reduction techniques to identify significant variables along with dropping insignificant ones. This was followed by running a logistic model to determine which factors impacted the admission process. Accuracy of the model was determined and other models like SVM, KNN, decision trees were also tried.

**Project 4: Querying in Hive for accessing various types of information from an employee data set :** Aim of the project was to write hive commands to access different types

of information from an employee rating data set from six major companies ( Amazon, Google, Apple, Facebook, Netflix and Microsoft). Data was provided for 67,000 employees. The specific steps in this project involved uploading the given data set using FTP (from simplilearn practice labs) onto hdfs, imputation by replacing the none entries in the various rating fields by the median values, table partitioning and bucketing of the table and creating new tables for satisfied and unsatisfied employees.

### **Technical Tools Known to me:**

Python, Tableau, Matplotlib, TensorFlow, Keras, Pandas, NumPy, Scikit-learn, SQL Query writing.

### **CERTIFICATIONS:**

❖ Awarded Certificate for completing online course on ‘Genomic Data Science with Galaxy’ authorized by Johns Hopkins University and offered through Coursera in Jan, 2021.

❖ **Certificate for Simplilearn’s Data Scientist Master’s program, in collaboration with IBM:**

Mastered the concepts of statistics, Python and R programming, machine learning, Hadoop ecosystem, and its deployment, and Spark. Gained hands-on exposure to organizing data, performing data analysis using R and Python, implementing machine learning models, building visualizations, and designing dashboards using Tableau. This program provided introduction to integrated blended learning. It gave the opportunity to apply skills learned by solving a real-world, industry-aligned data science problem as part of the capstone project. Successfully completed all the mandated course requirements and industry projects with distinction on August 22, 2022.

❖ **Certificate for Simplilearn’s Artificial Intelligence Engineer Master’s Program, in collaboration with IBM:**

Learned and mastered the concepts of Artificial Intelligence and its business applications, the fundamentals of statistics, Python programming, machine learning, deep learning, and natural language processing. Gained hands-on exposure to building artificial neural networks and traversing layers of data abstraction. This program gave an introduction to integrated blended learning while providing the opportunity to implement the skills learned by solving a real-world, industry-aligned Artificial Intelligence problem as part of the Capstone project. **Master’s Program ‘Artificial Intelligence Engineer’** offered by Simplilearn: Completed all the mandated course requirements and industry projects with distinction on January 23, 2023

❖ Awarded certificate of completion for successfully completing the OmicsLogic courses: “**Transcriptomics**” and “**Introduction to Bioinformatics**”. These courses involved completion of training and the understanding of concepts, requirements and

the logic behind bio informatics analysis as presented on the OmicsLogic training platform.

- ❖ Awarded certificate of completion for successfully completing the OmicsLogic course: **“Introduction to Bioinformatics”**.(June 2022)
- ❖ Awarded certificate for completing the course “ Getting started with Data Analytics on AWS”, an online course authorized by Amazon Web Services and offered through Coursera.(Oct. 2023)
- ❖ Awarded Udemy certificate of completion for the course “ Looker for Beginners”. (Nov. 2023).
- ❖ Awarded Udemy certificate of completion for the course “ Looker Studio/Google Data Studio Complete Advanced Tutorial”. (Dec. 2023).
- ❖ Awarded Udemy certificate of completion for the course “ GA4 Essentials Guide: Transitioning from UA to GA4”. (Nov. 2023).
- ❖ Awarded Udemy certificate for completion of the course “Intro to Big Data & Decisions in Healthcare, Plain and Simple. (May 2024).

## **PUBLICATIONS:**

- ❖ Tumor Suppressor Protein p53 Recruits Human Sin3B/HDAC1 Complex for Down-regulation of its Target Promoters in Response to Genotoxic Stress". Nidhi Bansal; Rama Kadamb; Shilpi Mittal; **Leena Vig**; Raisha Sharma; Bilikere S Dwarakanath; Daman Saluja, PhD. PLoS ONE pone.0026156.ISSN number : 1932-6203
- ❖ Vig, L. (2014) Comparative Analysis of Different Classifiers for the Wisconsin Breast Cancer Dataset. *Open Access Library Journal*, 1: e660.<http://dx.doi.org/10.4236/oalib.1100660>; ISSN no.2333-9721
- ❖ Contributed in the **e-PG Pathshala Project by UGC under NMEICT mission of Ministry of Human Resource Development (MHRD), Govt. of India**, as content writer (Co-Author) in Chemistry. Prepared 7 audio visual modules as part of MOOCs

