

# Roshankumar S

Data Scientist

[in www.linkedin.com/in/roshankumar-s](https://www.linkedin.com/in/roshankumar-s)

[✉ rkroshan00700@gmail.com](mailto:rkroshan00700@gmail.com)

[🌐 https://github.com/RoshankumarS14](https://github.com/RoshankumarS14)

[☎ +91 95144 38756](tel:+919514438756)

[📍 Bangalore, India](#)

## Profile

Highly motivated and dedicated Data Scientist with a passion for deciphering complex datasets to extract valuable insights and tackle challenging business problems. Proficient in leveraging statistical analysis and machine learning to drive data-driven solutions. Possessing a strong background in mathematics and programming, complemented by a relentless pursuit of excellence in the realm of data science.

## Technical skills

<b>Python</b>	Numpy, Pandas, Matplotlib, Seaborn, Scikit-learn, Scipy, HuggingFace
<b>MySQL</b>	Group By, Aggregate functions, Window functions, Joins, Subqueries
<b>Exploratory Data Analysis</b>	Feature Extraction, Feature engineering, Data preprocessing and wrangling, Scaling and Transformation, Dimensionality reduction, PCA
<b>Statistical techniques</b>	Hypothesis Testing – parametric and Non-parametric, Z-test, t-test, ANOVA test, Chi-Square test, Probability distributions functions
<b>Machine Learning</b>	Regression, Classification, Ensemble techniques, Clustering, Model optimization, Hyper parameter tuning, Cross validation
<b>Data Visualization</b>	Tableau, PowerBI
<b>Microsoft Excel</b>	Formatting, Functions, Sorting, Filtering, Pivot tables, Charts, Lookups
<b>Deep Learning</b>	Open AI, Transformer, TTS, RNN, LSTM, Lang Chain
<b>LLM</b>	NLP, TF-IDF, Bert, Sentimental Analysis, HuggingFace, Web Scrapping
<b>Cloud</b>	PySpark, Azure DataBricks, Data Lakes

## Work Experience

### Wipro Ltd

Feb 2022 – Sep 2022

.Net Full Stack Developer

- Learned Front End Development using Angular, HTML, JS, TypeScript, CSS
- Learned Back End Development using .Net core, EF core, Microservices and Microsoft SQL server
- Developed website for online book store using front end, back end and SQL server

## Real-time Project

### Voice cloning with Multilingual translation (Deep Learning)

- Leveraged ASR to transcribe spoken English audio into text, capturing every nuance of expression.
- Utilized language translator to convert English text into the enchanting script of Hindi, preserving the essence of meaning.
- Employed TTS to seamlessly translate English text into Hindi while retaining the original voice, demonstrating prowess in natural language translation and audio synthesis.

### Development of Personalized Recommendation System for Ecommerce (LLM/NLP)

- Developed a recommendation system using sentiment analysis through TF-IDF vectorization.
- Implemented a hybrid recommendation approach, combining collaborative and content-based filtering by creating product-product metrics and user-user metrics.
- Deployed the recommendation system on the website, for improving customer satisfaction.

### Car price Prediction using Regression and Ensemble techniques (Machine Learning)

- Employed Exploratory Data Analysis and Statistical analysis to gain insights.
- Applied Principal Component Analysis for feature reduction, enhancing model efficiency.
- Developed and evaluated multiple models (Linear Regression, Random Forest, Decision Tree, Ada Boost, Gradient Boost). Identified Random Forest as the optimal model with 90% accuracy.

## Internship

---

### Oasis Infobyte

July 2023

Data Science

- Developed a robust sales prediction model using advertising data, achieving an impressive R-square of 0.92. Employed advanced statistical techniques for accurate sales forecasts, optimizing decision-making.
- Created a highly accurate email spam detection model using Logistic Regression and TF-IDF vectorization, achieving an 96.42% accuracy. Skillfully categorized emails into spam and ham groups.

### KPMG

June 2023

Data Analytics

- Enabled company growth via data-driven insights and strategic planning. Ensured accurate data quality, extracted valuable insights, and presented findings visually for informed decisions.

### Lets Grow More - VIP

May 2023

Data Analytics

- Conducted in-depth stock price analysis via Time-series analysis. Employed stacked LSTM for stock prediction and forecasting, achieving an impressive accuracy of 98.75%.
- Performed thorough analysis on the iris flower dataset, revealing patterns. Employed Decision Tree for successful classification with 97.33% accuracy.

### The Sparks Foundation

Apr 2023

Data Science and Business Analytics

- Predicted student marks using Supervised Learning by developing a Linear Regression model with an accuracy of 92.91%.
- Predicted classes of Iris flowers using Unsupervised Learning by developing a K-means Clustering model with an accuracy of 90%.

## Education

---

### Post Graduate Program

Data Science and Engineering (Percentage: 90)

Great Learning

Sep 2022 – Present

### Bachelor of Technology

Chemical Engineering (Percentage: 84.5)

Sri Venkateswara College of Engineering

Aug 2018 – Jun 2022

### HSC

Percentage: 89

MLM Mamallan Mat.Hr.Sec.School

Jul 2017 – May 2018

### SSLC

Percentage: 94.8

MLM Mamallan Mat.Hr.Sec.School

Jul 2015 – May 2016

## Courses and Achievements

---

- Attained a Gold badge in Python by HackerRank.
- Attained a Gold badge in SQL by HackerRank.

## Soft Skills

---

### Analytical

Proficient in data analysis and deriving actionable insights.

### Creative Problem-solving

Demonstrated ability to approach challenges with a creative mindset, offering innovative solutions to complex problems.

### Communicating

Excellent verbal communication skills and written skills, able to convey information clearly and effectively.

### Presentation

Skilled in creating visually appealing presentations and deliver it with clarity.

### Risk Management

Active trader in Stock market with domain knowledge in Finance.

## Passion

---

Trading , Piano , Body Building , Badminton