

# Subhash Dixit

West Bengal, India

9205979486 | subhashdixit17@gmail.com | github.com/subhashdixit | linkedin.com/in/subhash-dixit-nit-durgapur

## Work Experience

### Affine

Bangalore, India

Senior Associate Data Scientist

Apr 2023 - Present

- Led a **3-member** team in scraping data from **100s** of pages across **6** gaming sites using **GCP OCR API, BeautifulSoup, Selenium, and Scrapy**. Created concise **dashboards** with **10+** charts for Weekly, Monthly, Seasonal, Sentiment & Battle Pass insights using Google Sheets.

### Affine

Bangalore, India

Associate Data Scientist

May 2022 - Mar 2023

- Supervised a data science team in the generation, segmentation, and clustering of a **40 KPI** analytical dataset of **23M players** from **10 tables** into **7 groups** with a **silhouette score** of **0.35** using **EDA clustering techniques** (K-means, factor analysis, elbow curve)
- Provided leadership with real-time visibility in compiling and predicting **retention & ARPPDAU** of **720 features x 180 rows** dataset for **720 days** using **FB-Prophet** with a **MAPE** of **10%** and **revenue improvement** of **2%**
  - Comp data was considered as input
  - Data was made **stationary** by EDA, pre-processing, and implementation of **time series models**
- Analyzed hot topic sentiments from web servers Reddit, Twitter, and Steam by **scraping 30k** comments, likes, topic headings, and descriptions using **RedditAPI, Praw API, tweepy, and BeautifulSoup** library in Python
  - Data was further analyzed and understood using three **NLP techniques, VADER & NGrams**
  - Build a UI interface using **CustomTkinter** and executable file to automate all the task using **PyInstaller**
- Designed **three live revenue models** in Excel on competition data for upcoming games to understand their marketing and incentivization strategies
- Designed a live dashboard on **Looker** with a **SQL** developer team for a game during its launch to assess its real-time performance, track its **key revenue** and **engagement metrics**, and to use such decisions for the company's ongoing databases
- Worked on research projects in gaming, performed **ad-hoc analysis**, and extracted relevant data by identifying **KPIs** from the telemetry source tables of the games **using Looker**

### iNeuron

Bangalore, India

Data Science Project Intern

Dec 2022 - Jan 2023

- Insurance Premium Prediction** | (Python, ML(Regression), MongoDB, AWS, Streamlit)
  - Conducted **EDA**, pre-processing, and checked **linearity, normality (JB test)**, and **autocorrelation (Durban Watson test)** on **1 M** insurance data from **UCI** repository for a **regression** model
  - Implemented various **regression** models selected **Xgboost** based on its **adjusted r-squared** of **0.90** & deployed the project on **AWS EC2** as a **web application** using **Streamlit**

 **GitHub**

### Tata Consultancy Services

Bangalore, India

Assistant System Engineer

Jun 2021 - May 2022

- Performed system analysis, documentation, testing, implementation & user support for platform transitions using **SQL & Unix**
- Conducted **log analysis** to troubleshoot errors and automated suggestions for recurring job applications during deployments
- Completed several **soft skills & technical training** on **Python, SQL & Unix** during **6 months** training period, winner of the training quiz

## Education

B.Tech In Civil Engineering, NIT Durgapur - CGPA (8.71 / 10)

Aug 2017 - Jul 2021

Full Stack Data Science Bootcamp, iNeuron

Jun 2022 - Present

## Key Projects

**WhatsApp Chat Analyzer** | (Python, Pandas, NumPy, NLP, Streamlit)

 **GitHub**

- Analyzed **WhatsApp** chat data, studying trends and **sentiment** using **VADER**, and applied **NMF** for **topic modelling**. Designed a **Streamlit web app** for interactive data exploration and deployment.

**Credit Card Fraud Detection** | (Python, TensorFlow, ML, ANN)

 **GitHub**

- Acquired a dataset of 100K transactions from the UCI repository, constructed a classification model using an ML algorithm to detect fraudulent transactions, achieving 95% accuracy and an FN (False Negative) count of 4, with Xgboost selected based on accuracy and FN metrics

**Stock Price Prediction** | (Python, TensorFlow, Keras, LSTM, YFinance)

 **GitHub**

- Pre-processed NSE financial data, applied technical & fundamental analysis (6 indicators, 5 ratios), built LSTM price predictor (test r-squared: 0.85), and created a web app for NSE stock diversification & analysis

## Skills & Coursework

<b>COURSEWORK</b>	Machine Learning   Probability & Statistics   Calculus   Stock Market   Linear Algebra   Linear Programming   Economics
<b>SKILLS</b>	Python   Machine Learning   SQL   NLP   Time Series Forecasting   Deep Learning   Statistics   Looker   Streamlit   Power BI   Web Scraping   Advance Excel   Basic (AWS, Flask, JavaScript, HTML, CSS, C++)
<b>Libraries</b>	NumPy   Pandas   TensorFlow   Keras   NLTK   Datetime   VAEX   Sk-learn   Seaborn   Matplotlib   CustomTinker   PyInstaller
<b>Soft Skills</b>	Time Management   Teamwork   Problem-solving   Documentation   Hardworking   Presentation   Flexibility

## Extracurricular Activities & Achievements

- Earned "Best Employee of the Month" award for an exemplary display of transcending any designated call of duty
- Served as the captain of the NIT Durgapur Cricket team at the national level in 2019