

##### Name: Samir Banerjee

##### Email: [smrbanerjee43@gmail.com](mailto:smrbanerjee43@gmail.com) Phone: +91-7715026369



**Profile Summary:**

* MCA with total nine years nine months experience in database programming. Includes more than two years of experience as a data scientist and data engineer using Python. I have strong technical/analytical skills and can lead and work as part of a team. I am looking for the role of a data scientist, engineer or analyst in multiple domains.

|  |  |  |
| --- | --- | --- |
| **Name of the company** | **Duration** | **Designation** |
| **Capgemini Technology Services India Limited** | **March 2022-Till date** | **Senior Consultant** |
| Anunta Technology Management Services Ltd. | April 2020-March 2022 | Senior SME |
| Gap Bridge Software Services Pvt. Ltd. | Feb 2019 – April 2020 | Senior Engineer |
| Clover Infotech | Mar 2017 –Feb 2019 | Senior Engineer |
| CredenTek Software & Consultancy Pvt. Ltd. | May 2016 – Mar 2017 | Senior Engineer |
| Comtel Consulting | Jul 2014 – May 2016 | Engineer |
| Protex Computers Pvt. Ltd. | Aug 2012 – Jul 2014 | Engineer |



* Good knowledge of database *Stored-procedure*, *Function*, *Package*, *Trigger*, *Queries* and *Materialized view.*
* Good knowledge of file handling operations, parsing of JSON objects, functions, sets, dictionaries and built in sequence types like *list*, *tuples, range and strings* in Python.
* Good knowledge on extraction, transformation and loading of JSON data from Apache Kafka to Elasticsearch index using Python.
* Good knowledge of *Linear Regression*, *Logistic Regression*, model creation, VIF, residual analysis, outlier analysis, measurement of RMSE and R-squared value and model application for prediction analysis.
* Good knowledge of Confluent Kafka and Elasticsearch.
* Good knowledge of *Python* on *Data-bricks* and web scraping on HTML, XML data sources using requisite *Python* libraries.
* Good knowledge of *Probability distribution*, *Hypothesis tests, Descriptive* *analytics*, splitting, training and testing of a dataset using requisite *Python* libraries.
* Certification on *Machine Learning* With *Python*.(*Certificate code:2432060*)
* *Certified as Oracle Database SQL Expert (IZ0-047, Testing ID: OC0989907), Oracle Certified Associate (IZ0-042, Testing ID: OC0989907), Oracle Certified Professional (IZ0-043, Testing ID: OC0989907).*
* *ITIL Foundation certification in IT Service Management (ITIL 4 edition).*

**Skills**

|  |
| --- |
| Machine Learning |
| Python |
| SQL |
| Data engineering |
| Data-bricks |
| Elastic search |
| PySpark |

**Academic Qualification**

|  |  |  |
| --- | --- | --- |
| **Degree** | **Year of Passing** | **Name of the Institution** |
| MCA | 2011 | NSOU-CMC |
| BCOM | 2002 | University Of Calcutta |
| Class XII | 1998 | CHSE(Bhubaneswar) |
| Class X | 1996 | ICSE |

**Project Experience in Machine Learning using Python:**

**Name: Retail Analysis with Walmart**

The aim of this project was to overcome the business challenge faced by Walmart, one of the leading retail stores in the US, due to unforeseen demands and out-of-stock issues. Walmart would like to predict the sales and demand accurately. Walmart runs several promotional markdown events throughout the year. These markdowns precede prominent holidays, the four largest of all, which are the Super Bowl, Labor Day, Thanksgiving, and Christmas.

An ideal ML algorithm will predict demand accurately and ingest factors like economic conditions including CPI, Unemployment Index as also the effects of markdowns on these holiday weeks.

Basic analytical tasks performed using *Pandas* and *Matplotlib* library included:

* To find stores with maximum sales.
* To find out which store had maximum standard deviation i.e., the sales vary a lot. Also, find out the coefficient of mean to standard deviation.
* To find out which stores had a good quarterly growth rate in Q3’2012.
* To find out holidays which have higher sales than the mean sales in non-holiday season for all stores together.
* To provide a monthly and semester view of sales in units.

For building, training and testing of Prediction Models *Scikit-Learn, Matplotlib and Statsmodel libraries were used.* Tasks included:

* Linear Regression – Analyze whether CPI, unemployment and fuel prices have any impact on sales.
* Change dates into days by creating new variables.

Project at **Capgemini Technology Services India Limited**

|  |  |
| --- | --- |
| **Client** | Chubb Inc. |
| **Project** | Identity Management |
| **Environment** | Python 3.9, Visual Studio, Apache Kafka, Elasticsearch, Linux |
| **Position held** | Senior consultant |
| **Duration** | April 2022- July 2022 |

**Responsibilities:**

* Responsible for coding the data ingestion pipeline in Python from Apache Kafka to Elasticsearch as part of the Identity Management project.
* The pipeline extracts data from Apache Kafka to a staging index and then from the staging index to the reporting index in Elasticsearch. The data is then used in the Kibana dashboards for visualization.
* The pipeline code is written in Python and extracts raw data from Apache Kafka to the Elasticsearch staging index in the first stage. In the second stage it applies transformations to the data as required before pushing into the reporting index.
* Uploaded the script, configuration and other parameter files in the Git repository and cloned the repository in Git Bash.
* Created a Cron job that is scheduled to execute the script after every 10 min and log the execution steps in a log file.
* Update the Jira dashboard and provide updates to the scrum master on the day’s work.

**Project at Anunta Technology Management Services**

|  |  |
| --- | --- |
| **Client** | LTFS |
| **Project** | Investment Management. |
| **Environment** | Python 3.9, Anaconda 3, Jupyter, Oracle 11g, SQL Server 2012. |
| **Position held** | Team lead |
| **Duration** | 2020 – 2022. |

**Responsibilities:**

* I lead a team of four people.
* Do data extraction and normalization.
* Create data frames and split into training and test datasets.
* Build Regression models on the training dataset.
* Validate the model using residual and outlier analysis.
* Apply the model on the test data to predict ROI, product wise sales, revenue growth, customer base, customer segment, market share and other KPI. Model used to create business rules and MIS dashboards.
* Assist the team in coding SQL queries and database packages when required.
* Attend meetings to discuss pending issues with higher verticals.

**Project at Gap Bridge Software Services**

|  |  |
| --- | --- |
| **Client** | LTFS |
| **Project** | Investment Management. |
| **Environment** | Oracle 11g, SQL Server 2012, Linux (Red Hat 7.2), Windows 10. |
| **Position held** | Senior Engineer |
| **Duration** | Feb 2019 – April 2020 |

**Responsibilities:**

* Meet technical challenges arising out of database or application downtime, middleware issue, network issue, database or application slowness.
* Write queries, package and SQL job, debugging, query optimization and tuning.
  + - * Data load with Oracle loader.
      * Attend meetings to discuss issues pending and expectations from the client side.

**Project at Clover Infotech**

|  |  |
| --- | --- |
| **Client** | LTFS |
| **Project** | Investment Management. |
| **Environment** | Oracle 11g, SQL Server2012, Linux (Red Hat 7.2), Windows 10. |
| **Position held** | Senior Engineer. |
| **Duration** | Mar 2017 – Jan 2019 |

**Responsibilities:**

* + - * Involved in project migration. Tested and resolved connectivity and login issues on multiple, new databases, middleware issues, recompilation and testing of database objects, query optimization and tuning.
* Create queries, package, SQL job, debugging, scheduling and automation of recurring activities.
* Data load with Oracle loader.

**Project at CredenTek**

|  |  |
| --- | --- |
| **Client** | Family Credit |
| **Project** | *Retail lending* |
| **Environment** | Oracle 10g, SQL Server2008, Linux (Red Hat 6.2), Windows 7. |
| **Position held** | Senior Engineer. |
| **Duration** | 2016 - 2017 |

**Responsibilities:**

* Create query, package, *SQL* job and do debugging.
* Data load with *Oracle* loader.
* Tuning the queries and packages for optimal performance.
* Create scheduled jobs for generating daily reports.

**Project at Comtel**

|  |  |
| --- | --- |
| **Client** | Bajaj Finserv |
| **Environment** | Oracle 10g, SQL Server2008, Linux (Ubuntu 6.06), Windows 7. |
| **Position held** | Engineer |
| **Duration** | 2014-2016 |

**Responsibilities:**

* Create query, package, *SQL* job and do debugging.
* Data load with *Oracle* loader.
* Tuning the queries and packages for optimal performance.
* Create scheduled jobs for generating daily reports.

**Project at Protex**

|  |  |
| --- | --- |
| **Client** | Coal India |
| **Project** | Coalnet |
| **Environment** | Oracle 10g, Sun Solaris, Java-script 6.0. , Windows. |
| **Position held** | Engineer |
| **Duration** | 2012-2014 |

**Responsibilities:**

* Created queries on Oracle production database for Payroll & PIS (Personnel Information System).Data provided information to the client related to salary, increment, leave, retirement, transfer, promotion and resignation of employees.
* Created and deployed reports at server level using Oracle Middleware tools like Report-Builder 6.5.
* Modified or enhanced existing Java Script pages at server level using tools like Oracle J-Builder and J-Developer.

**Personal Information**

|  |  |
| --- | --- |
| **Name** | **Samir Banerjee** |
| **Date-of-birth** | 08-07-1980 |
| **Passport No.** | P5912593 |

https://in.linkedin.com/in/samir-banerjee-08142425

https://r2d2.capgemini.com/IWRetainWeb/IWISAPIRedirect.dll/Files/static/html/webportal.html?page=talentprofile&resid=1631068

**---------------------------------------------------**

**(**Samir Banerjee**)**