|  |  |  |
| --- | --- | --- |
| Name: **Prosenjit Das** |  | Lead Data Engineer |
| |  | | --- | | Summary –13+ years’ experience in Software Development **Expertise** in Building Data Pipeline, ETL and Visualization  ***Contact details:*** *Sunnyvale, California- 94086, United State*  ***Mobile:****+1-669-264-0843*  ***Email-ID:*** *das.prosenjit.eie@gmail.com*  ***Linked-in****: https://www.linkedin.com/in/prosenjitdataanalyst* | | Technical Skill set | | **Data Engineering & Analysis**:   * Python, Java, Scala, Spring, Spark, Kafka, TensorFlow * Splunk, Github   **Data Visualization**:   * Tableau   **ML Algorithms :**   * K-Means, Random Forest, Classification, SVM, Collaborative Filtering   **Database**:   * Oracle, ElasticSearch, MongoDB * Customer Relationship Management – SugarCRM * 5+ years Data Engineering * 5+years Mobile Development * 2+ years Core App Services | |  | Currently Working for **Apple** (on contract basis), Sunnyvale, CA. At a high level-  Data Engineering:   * Experienced in Data **Collection**, **Filter, Integration, Analysis.** * Text Extraction from Legacy Data Sources and clustered into Structured Data. * Build pipeline through **Kafka, Spark Streaming** and **Microservice** * Built closed to Real-time data streaming by **Spark** with Parallel and Distributed Data process. * **Data lake** to hold RAW data, which help to do analysis and reusing for creating Weekly Forecasting Report. * Graphics Representation of different Reports in Tableau, Splunk   Data Analytics:   * Analysis on top of **Oracle** Database. Build Report on process status and failures. Data Validation Report, Performance Tuning. * **K-Means** (MLlib) to handle Huge Big Data Analysis on fly - Cluster / categorized data and then, push to different applications. * Used **Random Forest, Logistic Regression** for decision makingin **Tensorflow** and improved by Gradient Descent for solving complex problem on run time. * **NLP** on Invoice and user comments. Categorization based on content. Calculate and Build Excel, which reduce the manual effort for Sales Team. * Build **Recommendation** system by **Spark Collaborative Filtering** to help building dashboard for B2B users (Find similar users and Analyzed their history preferences). – Implicit and Explicit * Match and merge (inspired by Fraud Detection) on top of streaming data for grouping transaction. * Build **Chat-Bot** – Communicate with Database by local language   Performance:   * Improve the speed by Spark and Kafka tuning, SQL Indexing * Introduce **Multi-Threading** and refactor with cutting edge * Improve Cache by Elastic Search API for Search and Data Buffer * Persist the model and retrained with new data * Re-Architecture Data Pipeline and build all calculations on fly   Functional:   * Expertise to solve complex Business Problem * Build mapping between Non-structured data to Structural Components * Understand CRM lifecycle and build Business Intelligence on Enterprise Level |

**Work Experience**

1. **Apple** (on contract via Infosys)(April, 2014 – till date)

**Projects: CRM Platform -** Build Data pipeline to communicate with multiple platforms. Live and legacy system data has been captured and converted to usable CRM components. CRM is in central, which make Advantage for communicate with all other platforms and databases. Purpose to build smartest CRM to replace all existing CRM platforms.

* + Built Security and Performance for Streamed Data and make data available to CRM without fail (By Fault-tolerance, Cache multiple steps , Train existing Models)
  + Streaming
    - Categorized stream data and distribute to different data-set.
    - To improve performance, cache the data in Kafka and Elastic Search and reuse and expire as not required.
    - Mlib based Regression and classification library for analyse hidden fault
  + Recommendation
    - Based on Preference and Ratings of Product and Customer
    - Bring out the history and data exist in different platform.
    - Classified data on user comment and invoice
    - Used Random forest for decision making to identify quality of customer and rank them
  + Chat-bot
    - to help user directly communicate with Database and other platforms.
    - Reinforcement to learn by user response and improve bot service.
  + Decision Making
    - Used Random forest for decision making to identify quality of customer and rank them
  + Wrote Spring Services to communicate with Other Java Platforms.
  + Migrate
    - Old Architecture from Generic Queue-Java-SpringBoot to Kafka-Scala-Spark Platform to handle Huge Streaming Data coming from Legacy system and Realtime.
  + Data Lake
    - Keep on increasing data volume with increasing micro-services and platforms
    - Handle the data load tune the architecture.
  + to hold incremental data and further process for data analytics.
  + Dashboard
    - To get Splunk Performance graph depends on Service logs. Monitor Data lake and service failures.
    - Weekly Forcasting and Report for senior managers
    - Initial data validation and POC to find best algorithms
  + Worked on Oracle Report for Data Failure Analysis.

**Technology: Spring , Java, Spark-scala, Kafka , MLlib, Splunk, Github , Oracle, Elastic Search**

* **Product Tracker (**for Inventory Management system)
  + Build inventory management platform for Newly launched products. User can track the inventory details (like: location and delivery date) by map.
  + Send notification on registered devices.
  + Used “nearest neighbor” ML library to find nearest location of stores.
  + Linear regression to find cost of transport and packing
  + Analysed most effective centers and reliable demo happened in past.

**Technology: AngularJS, GoogleMap, IPhoneSwift , Tensorflow**

1. **Infosys Pvt Ltd.** (May, 2013 – till date) - https://www.infosys.com/

* **Micro-service** **Management**: Build Dashboard to automate Custom Services without human interference.
  + Automation build on top of Netflix Hystrix platform
  + MongoDB to store Service generated data.
  + Build UI to create Rule, which solve complex Service dependency layer and Services link to Each other.
  + Not effecting Legacy Service structure
  + Monitor the services (tracked in graph) and Auto Recovery, means never fail strategy.

**Technology: Java, Hystrix, Python**

1. **Hewlett-Packard Laboratory** – Services Global Delivery India(June, 2011 – May, 2013)(http://www.hpl.hp.com/)

**Role:** Data Analyst and Mobile Application Developer.

**Projects**:

* Advertise Suggestion Box(Recommendation System): Recognize the product from ongoing Videos and suggest similar products from web stores.
* Wrote algorithm to detect products like (shirts, cap etc.) from preprocessed videos.
* Tagged specific screens in videos and collect images. Process those images to find out product and compare the image by trained dataset
* Webstore API provide active Product price and details on the bottom of videos
* **Technology:** Java OpenCV, ML, Video Stream, Android Tablet ( 4.0)
* Crowd Cloud: Build the platform on top of Amazon Mechanical Turk in Handset Devices. Share Work like Video Recognition, Photo identification, Hand writing and Translation.
* Split the video in small pieces and trimmed same screens and process for Share with crowd
* Find out Images, which don’t have any objects and trimmed from image list
* Face cognition from videos ad images
* Recognise special words from scanned documents
* **Technology:** Java OpenCV, Face Recognition, Android Framework ( 4.0)

**4. July Systems (**http://julysystems.com/**),** **Bangalore (August 2009 – May 2011)**

**Project**: Live Gaming Application for Ongoing US based Sport Websites, Online Mobile Application for News and Adventure Channels.

**Technology Used:** Android Framework (ver 2.2), IOS Framework (Objective C++), Blackberry , HTML5, JavaScript, Google Maps API

**Role:** WAP and Mobile Application Developer.

Developed mobile application for live sport sites

**Education:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Graduation** | **Year** | **University** | **Grade** |
| B.Tech | 2006 | West Bengal University Of Technology (**W.B.U.T**), India | 78.3% |

**Certificates:**

* MongoDB for NodeJS from MongoDB University
* Machine Learning from Stanford University (Coursera)

**General Information:**

**Passport:** Valid up-to 2026

**Visa Status:** H1-B (started from December, 2015)

**Date of Birth:** 29th May, 1983

**Born in:** Kolkata, India