Sandeep Singh

+91-8750227690

sndp1811@gmail.com

Data Scientist | Machine Learning Engineer | Technical Architect | Scrum Product Owner

Summary

- **IIM - Calcutta** certified Data Science & Big Data Analytics professional with **7+ yrs' experience** implementing Machine Learning across Manufacturing, E-commerce, Retail, Finance, Automobile, HRM Industrial domains.
- Engineered over **10+ enterprise-grade** Data Science products, consisting of **Predictive Modelling**, **Chatbots**, **Forecasting**, **NLP**, **Computer Vision** & **Deep Learning** Applications, and Business Intelligence **Dashboards**.
- Spearheaded 5 developers as **Tech Lead** while also contributing individually as a **Full Stack Machine Learning** Engineer. Proficient in designing data pipelines for both **SQL/NoSQL**, training ML-Deep Learning models using **GPU-NVIDIA CUDA** architecture, pip installable **packaging**, designing **REST APIs** and **Microservice Architecture**, deploying with **Docker, AWS EC2**, and **Apache HTTP Web Server**.
- Expertise in **MLOps, TOGAF, JIRA, Agile-Scrum** and **CMMI Level-5** software development methodology.

Skills

Python, Packages & Libraries

Data Types, Operators, Control Flow, Functions, OOPs, Errors and Exceptions, Modules & Custom Packages, Input and Output, File Handling, Standard Libraries, Logging, Multiprocessing and Multithreading, Pandas, NumPy, Matplotlib, stats models, Scikit-Learn, PyCaret, Flaml, Keras, TensorFlow, PyTorch, spaCy, Fast Text, Feature-Engine, mlflow, Weights & Biases, Django, sqlalchemy, OpenCV, YOLO, GPU-NVIDIA CUDA Toolkit

SQL

Querying data, filter data, Group By, Joins, Subqueries, Set operators, Modifying data, Managing databases & Tables, Data types, Constraints, Views, Indexes and Triggers

Data Visualization

Bar Charts, Heatmaps, Histograms, Scatterplots, Boxplots and Line plots

Machine Learning & Deep Learning Algorithms

Regression, Tree Based Modelling, Gradient Boosting, Bayes' Theorem, Clustering, SVM, Time-Series, Bias & Variance, Regularization, Dimension Reduction, Cross-Validation, Parameters Tuning, LSTM-RNN, CNN, BERT, GPT-2, Gradient Descent Optimization, Word Embedding, NER, OCR, Text Classification, Language Modelling, Search Engine, Hugging Face Transformers, Object Detection, Chatbots, Topic Modelling, Sentiment Analysis

Probability & Statistical Techniques

EDA, Descriptive Stats, Probability Distributions, Inferential Statistics, Hypothesis Testing, Feature Engineering, Accuracy Metrics, Bayes Theorem, Central Limit Theorem, Univariate & Multivariate analysis

MLOps & Agile Project Management

Git, SVN, Python Package, Django REST API, Docker, CI/CD - Jenkins, Weights & Biases, MLflow, SonarQube, Postman, Linux, Apache Web Server, Azure, AWS - S3 bucket, RDS, Elastic Compute Cloud (EC2), Elastic Container Registry & Service (ECR), Azure DevOps, MySQL, MongoDB, PayPal payment gateway

Education

- **PG Diploma:** Data Science & Big Data Analytics – **IIM Calcutta,** 2017-18, A-Grade
- **PG Diploma:** Applied Statistics with specialization in Industrial Statistics, 2016-17, 1stDivision
- **Diploma:** Computer Applications and Programming, 2010-11, A+ Grade
- **B.Sc.(H) Physics: University of Delhi**, 2009-12

Experience

- **Nityo Tech:** Tech Lead – AI ML & Data Science — Apr 23 – Present
- **1CloudHub:** Tech Lead – AI ML & Data Science — Aug 21 – Mar 23
- **Nityo Infotech:** Sr. Data Scientist | Scrum Product Owner — July 20 – July 21
- **Minda Industries Ltd:** Assistant Manager – Data Scientist — Mar 18 – July 20
- **US Tech Solutions:** Data Analyst — Feb 15 – April 17

Projects

1. Predictive Modelling for a B2C Mobile App platform:

- **Classification and Regression models** were developed for a digital marketplace platform utilizing the user's identity data, companies offer data, marketing campaign and financial transactions data to:

a. **Predict the probability** that the concerned user will be a consumer of a certain company's products/services.

b. **Predict the amount** they are estimated to spend on the company's products or services in a fixed duration.

- Some of the data preparation steps included missing data imputation, Feature Scaling to the median and quantiles, creating dummy variables, feature selection for reducing correlated variables and feature space, combining multiple features to create new features, and creating new datetime features.
- **Regression, Logistics, Random Forest, XGBoost,** K-fold cross-validation algorithms were utilized to develop and achieve more than 75% accuracy on both the models.
- Integrated trained ML Models into **Mobile app** using **REST APIs** utilizing Django REST framework.
  URL: https://play.google.com/store/apps/details?id=com.nityo.identitywallet

2. NLP based Search/Matching Engine using Word embedding Techniques:

- The project involved developing a **Matching Engine** to compare **Job Description** docs with stored **Resumes** documents in MongoDB database using the **Matrix Similarity algorithm**. The documents were **ranked** based on their **similarity scores** in descending order, with the highest-ranked document being the most similar to JD.
- Data pipelines were developed to fetch, and clean resumes text stored in a **MongoDB database** using a variety of techniques such as converting to lowercase, removing punctuation, stop words, and whitespace etc.
- Cleaned text then transformed into word vectors using multiple **word embedding** techniques like **Word2Vec, Doc2Vec and FastText**, and trained **Deep Learning model** on vector corpus and serialized using **joblib** module.
- Trained ML model integrated into **REST API**, that takes **base64** encoded JD text, perform a **similarity query** by fetching CVs corpus from MongoDB collection, sort the resumes by **similarity score**, and created JSON output.
  URL: https://livedeeplearningparser.mycareercube.com/matching_engine/

3. Conversational AI/ Chatbot software for building text and voice-based assistants:

- Utilized Python and the **RASA bot framework** to create Chatbots for various applications, including **lead generation** and bots specifically designed for the **HR** and **Finance** domains.
- Integrated **ChatGPT** & other APIs, **email trigger** functionality, **role-based** access control, and used **MySQL/MongoDB** databases with a Python Action Server for real-time data retrieval via text/voice queries.
- Articulated Intents, Rules and Stories and Integrated **Slot**, **Forms** with Chatbot NLU, Machine Learning engine.
- Deployed and Integrated bot on **websites** and **WhatsApp** platform using REST, Twilio and Socket-IO channels.
  URL: https://nityo.com/

4. HR Recruitment Automation using Deep-Learning, NLP, Optical Character Recognition(OCR):

- **NER Label prediction in JDs and CVs Docs** – NER based parsing engine for extracting structured labels related to job, experience, skills, personal info etc from the unstructured input text data.
- Pre-processed, Analysed and Labelled unstructured text data of more than 20K JD/Resume documents and devised more than **50^{+} NER labels** using active Learning based **Prodigy data Annotation** tool.
- Integrated data pipeline with **AWS RDS & S3 Bucket** for storing, retrieving, and training deep learning models.
- Trained/Finetuned custom **spaCy NER** and **Transformers** model for Named Entity Recognition, on **NVIDIA GPU – CUDA** based architecture on **AWS server**.
- Converted both Deep Learning models into **pip installable python packages** & integrated into REST APIs.
- Implemented **base64**, **JWT** Bearer Tokens, **HTTP** response status codes, body parameters, Optical Character Recognition (**OCR**) functionality using **Django REST APIs** for reading & parsing Image, PDF & Word documents.
  URL: https://ai.mycareercube.com/admin/upload-resume-new

5. Organization's CMMI-5 Appraisal review | Agile Scrum Implementation:

Appraisal Details: https://www.cmmiinstitute.com/pars/appraisals/58533/details

- Played a pivotal role in achieving the highest possible *CMMI-5 appraisal rating* for the organization by leading the Tech Team as the ***Scrum Product Owner*** utilizing **Agile Scrum development lifecycle**.
- Utilized techniques including, project planning, backlog management, sprint calculation, estimation, metrics design, communication planning, and risk management to ensure project success and enhance team

efficiency, resulting in a **25% reduction** in both project costs & project completion time.

6. Python based Automation:

- **File Format Conversion:** Bash script to convert large volumes of PDF and Word document files from a local system into JSON, JSONL file format using glob, OS, JSON, PyPDF2, pytesseract, pillow & file handling modules.
- **Cron Job** for **storing word embedding data in MongoDB database** using joblib, pymongo, pandas and Gensim modules, developed using a variety of techniques including Word2Vec, Doc2Vec, Fast Text, and LdaModels.
- **Cron Job** for **Batch/bulk prediction** of **Text Classification model** and storing the result in the MySQL database.
- **Auto Email Triggers:** Cron Job for sending daily basis multiple email triggers to cross functional team members reminding them to do their planned financial assignments by selecting pending tasks from a MySQL database and ensuring they are completed by the due date using smtplib, SQL alchemy, datetime, JSON modules.
- **Automate SVN commits and updates** on AWS EC2 and Apache HTTP Server using a Bash script that can be executed on VS Code Server.

7. Supply Chain Automation using Demand Forecasting and Inventory Management :

- Implementing SKU wise **demand forecasting model** using historical Sales data of **Automobile domain** including Order Booking, Discount schemes, Pricing, and Sales Promotional Activities data.
- Analysed, Cleaned and Pre-process data and Engineered data pipelines from SAP HANA, Infor BAAN ERP, Web-portals, and SQL Server databases, for the continuous flow of more than **4000 SKUs data**.
- Devised Regression and Time-series forecasting for developing model ML Model, resulting in an SKU-wise accuracy of more than 80%.
- Implemented **Theory of Constraints** for developing inventory levels, automatically trigger reorder points, and optimize the allocation of stock across multiple locations by considering factors like lead times, customer demand, and production schedules.
- Yielded a **cost savings** of nearly Rs **6 crores** by reducing order losses from 15% to 5% of its monthly sales for an 800-Crores Annual Turnover Business.

8. Real-time Business Intelligence Analysis and Automation using Tableau Dashboards:

- Developed real-time Business Intelligence **Tableau dashboards** across business verticals and functional domains such as Sales & Channel Marketing, Product Marketing, Finance, HR, and Logistics that assisted in real-time data-driven decision making and amplified total business **profitability by 10%.**
- Automated MIS reports by creating data pipelines, organizing, compiling, and querying data from SQL Server and SAP HANA databases.
- Formulated Tableau **Server Admin**, Extract **Refresh Schedules**, Users & Projects **Permissions**, maintained server up-time by 99% and authorized more than 500 Tableau licenses across top & middle management.

Training and Certifications

- Becoming An Effective Trainer (TTT) by Mercury Goldmann
- Change Management
- Supply Chain Management by Maruti Centre for Excellence
- Business Finance by Tata Steel Management Development Centre
- ISO/IES 27001 Information Security Management
- Enterprise Architecture - TOGAF 9.2

Links

- https://www.linkedin.com/in/sandeep-singh-iimc/
- https://github.com/Sandy1811
- https://www.hackerrank.com/profile/sndp1811