PUSHPENDU DAS

Senior Data Scientist

Contact No: +91 9916412192
E-mail Id: pushpendu.das1112@gmail.com

GitHub   Linkedin
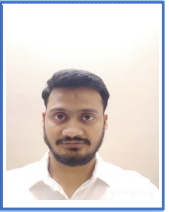
PROFESSIONAL SUMMARY:

- ➢ Senior Data Scientist with total 7+ years of experience in IT Industry.
- ➢ 5+ years' experience in Data Science, 2 years of experience on cloud IaaS and PaaS service with strong background in database.
- ➢ As a programming language, I have worked on **Python** extensively and hold good knowledge of implementing OOPs concept.
- ➢ Relevant knowledge of supervised and unsupervised learning (like **Regression, Classification, Ensemble, Clustering**).
- ➢ Worked on **end-to-end ML pipeline** and deployed on GCP using **Kubeflow**, **Docker** as **MLOps** technique.
- ➢ Good experience in **Natural Language Processing** using NLTK, Spacy and Natural Language Understanding using **Transformer**, **BERT**.
- ➢ Having good experience in using Word Embedding technique like TFIDF, **Word2Vec**, Glove, Elmo etc.
- ➢ Having good knowledge in using DL algorithm like ANN, RNN, **LSTM**, CNN, Encode Decoder, **Seq2Seq**.
- ➢ Work on several NLP task like **Text Classification**, Question Answering, Information Extraction, Text Summarization, **Keyword Generation**.
- ➢ Work on several CV task like Image Classification, **Object Detection**, **OCR**, Color Detection.

Technical Skills:
- ➢ Data Science Skill   : Machine Learning, Natural Language Processing, Deep Learning, Computer Vision
- ➢ Cloud Technology   : Azure IaaS, AWS IaaS, GCP
- ➢ Scripting Language   : Python
- ➢ Data Science framework   : Numpy, Pandas, Sklearn, Spacy, Pytorch, Transformer, YOLOv3, Amazon Rekognition, Vertex AI
- ➢ MLOps   : Docker, Kubeflow Pipeline
- ➢ Database   : MySQL, PostgreSQL, Big Query
- ➢ IDE   : Anaconda, PyCharm

PROFESSIONAL EXPERIENCE

Employer: Sabre Corporation
November 2021 to present, Bangalore
Senior Data Scientist

Project: Customer Behavior Forecaster

Process overview: Company wants predict the probability of selecting itinerary by customer which help to generate automated dynamic travel packages based on origin, destination, travel date etc. This will reduce customer trip planning effort and generate revenue.
Roles: Senior Data Scientist
Responsibilities:
- ❖ Built a tensorflow **time distributed LSTM** model using keras to predict probability of itinerary selection with **90%** accuracy.
- ❖ Achieved serverless deployment using **dockerization**, **kubeflow** pipeline via CICD Pipeline and **github** version control.
- ❖ Managed GCP resources using **terraform** script, handled docker image using **bazel build** and testing the code using **pytest**.

Tools and technology: Python, **Tensorflow**, **LSTM**, Pytest, Terraform, **Docker**, **Kubeflow**, GCP, Git, Bazel.

Project: Market Segmentation

Process overview: Market segmentation engine will try to cluster the similar type of travelers, which will help to analyze the trend of trip. Customer will be able to understand behavior of market, movement of travel in respect of day, month and year.
Roles: Senior Data Scientist
Responsibilities:
- ❖ Extracted MIDT rawdata from GCP Bucket, and preprocessed the data to make structure tabular format then stored in **Big Query**.
- ❖ Built Big Query **XGBoost** classification model to identify trip type and **Kmeans** cluster model to create the market segment.
- ❖ Deployed the process on **GCP** through CICD Pipeline with **simple compute** technique, GCP Resources handle by **terraform** script.

Tools and technology: Python, GCP, **Big Query**, **XGBooast**, Kmeans, Pytest, **Terraform,** Jenkins,

Employer: Ideapoke Technologies Pvt. Ltd.
April 2020 to November 2021, Bangalore
Lead Data Scientist

Project: Optimized Search Engine for Technology Scouting

Process overview: Optimized search engine has tried to bring compact information instead of proving generalized information, where requirement is about technology scouting information. Which contribute information about start-up companies, research, patent etc.
Roles: Senior Data Scientist
Responsibilities:
- ❖ Supporting billions of data and insight from the data like summary, important keywords, related technology etc.
- ❖ Increased the productivity of analyst **by 90%** and **search time** for customer has reduced **10 times** than before
- ❖ Build automated keyword extraction and automated text summarization model, accuracy for both models nearly **84.6%**

Tools and technology: Python, **NERS**, TF-IDF, custom Word2Vec, ELMO, Text Ranking, Linear SVC, **LSTM**

Project: Automated Market Scouting

Process overview: Worked on a Module called AI-powered data. Odyssey product used as a platform for a customer, where user can post their requirements regarding market scouting information like market size, market trend, technology trend, strategy etc.

Roles: Lead Data Scientist

Responsibilities:
- ❖ This project has increased **productivity** of the company **till 60%** and **time** has reduced to **4 times** than before
- ❖ Used of QnA model for contextual information extraction from co-referenced text based on Fine Tuned BERT model
- ❖ Text classification with word embedding technique like BERT tokenizer and ML model like Linear SVC

Tools and technology: Python, Counter Vectorizer, TFIDF, SVM, **BERT classifier**, AllenNLP, Spacy, **CatBoost**

Project: Graph Data Extraction

Process overview: One of our critical tasks is information extraction from PDF related files. This issue came wherever information extraction is required from PDF. We found most of the data are in PDF but not able extract data from graph in structured format.

Roles: Senior Data Scientist

Responsibilities:
- ❖ Design and Implemented **YOLOv3** to detect graphs and graph component with **94% accuracy**, used GPU enabled machine till 14K epochs
- ❖ Developed image classification model using **resnet50+inception** model with more than **90% accuracy**
- ❖ Used AWS Rekognition **OCR** technique to extract the text from image color detection model using OpenCV with **96% accuracy**.

Tools and technology: Python, pdf2image, OpenCV, YOLOV3, resnet50+inception, Keras-OCR, AWS Rekognition

Employer: Racetrack.ai

July 2019 to March 2020, Bangalore

Software Developer

Project: Marvin Personalized Recommender System

Process overview: Built personalized recommender system using ML algorithm, NLP technique and recommendation technique. With help of personality, interest and user's details able to resolve cold start and new user preference problem.

Roles: Data Scientist

Responsibilities:
- ❖ Performed NERs, POS tagging for **text normalization** and built **recommendation engine**.
- ❖ Trained and performed hyper-parameter tuning SMV classification model to predict personality and interest with **85% accuracy**

Tools and technology: Python, TF-IDF, Linear SVC, NLTK, Spacy NER, IBM-translator

Employer: Tata Consultancy Service

December 2015 to July 2019, Bangalore

System Engineer

Project: Real time Customer Data Analytics Dashboard

Process overview: Developed advanced analysis, model driven, Customer sentiment and tone analytical dashboard to understand customer activity in better way and improve customer value in respect of needs.

Role: Junior Data Scientist

Responsibilities:
- ❖ Exploratory **data analysis** to provide understanding of leads and **word-cloud** provides most trending/search keywords
- ❖ Deployed predictive **classification model** like sentiment and tone analysis which help to increase the sells **around 30%**

Tools and technology: Python, Numpy, Pandas, Sci-Kit Learn, word cloud, NLTK

Project: Office Depot Optumera ADM

Process overview: Optumera has artificial intelligence module to align with customer expectations and enhance performance.

Role: Junior Data Scientist

Responsibilities:
- ❖ Collected unstructured and semi-structured data performed **exploratory data analysis** on product price and company profit
- ❖ Executed **market basket analysis** using association rule mining technique to find relevant paired products

Tools and technology: Python, Numpy, Pandas, Sci-Kit Learn, Seaborn, Matplotlib

COURSE AND CERTIFICATION:
- ✓ Learning Python for Data Analysis and Visualization
- ✓ Machine Learning A-Z™: Hands-On Python & R in Data Science
- ✓ Data Science: Natural Language Processing (NLP) in Python

EDUCATION

Passing Year	Degree	College	University
2015	Bachelor of Technology	Heritage Institute of Technology	Maulana Abul Kalam Azad University of Technology
2011	Higher Secondary	Midnapore Town School	West Bengal Council of Higher Secondary Education
2009	Secondary	Basantia High School	West Bengal Board of Secondary Education

PERSONAL DETAILS
- ➢ Date of Birth: 4th Aug 1992
- ➢ Gender & Status: Male & Single
- ➢ Languages Known: English, Hindi, Bengali
- ➢ Hobbies: Sports, Sci-fi movies, Travelling

EXTRA CURRICULAR ACIVITIES
- ➢ Member of University Sports Council
- ➢ Won medals in athletic games at college level competition
- ➢ Played Cricket, Football at college level