Priyanshu Shekhar Sinha

Bangalore | +91 7349575896 | priyanshu1996@hotmail.com | LinkedIn| Portfolio | GitHub | HackerRank

Profile

Data Scientist with 2+ years of data based product experience. Proficient in predicting modeling, data processing and data visualization as well as scripting language like python. Capable of creating, developing, testing and deploying data-driven services to translate business and functional qualifications into substantial deliverables.

Experience

DATA ANALYST | RECRUITMENT SMART TECHNOLOGY LTD PVT | MAY 21 – PRESENT | 5 MONTHS

- Designed a scalable solution to generate periodically opportunity report, operational report and advanced analytics report on the product performance for different clients.
- Reduced time for new client integration from 10 days (manual) to 3 hours i.e. 97% improvement.
- Reduced time for generating operational report (5 days) and opportunity report (1 day) to 30 minutes i.e. 95% improvement.
- Reduced time for generating advanced analytics report from 7 days to 15 minutes i.e. 99% improvement.
- Automated the Power BI dashboard for regular data updation.
- Technology Stack: Python, Data Analysis, Machine Learning, AWS S3, EC2, SES, Power BI, Docker

DATA SCIENTIST | SUMYAG DATASCIENCE PVT LTD| AUG 19 – APR 21 | 1 YEAR 9 MONTHS

- Designed and implemented an end-to-end scalable solution for extracting information from unstructured documents for enterprise clients.
- Achieved cross-domain compatibility by designing a highly configurable data-driven pipeline for domains like commercial insurance, healthcare, automobile, real estate and telecommunication.
- Leveraged AWS auto-scaled infrastructure to build custom statistical machine learning models for generating embeddings to predict entities from a corpus of 60000+ documents, roughly summing up to a terabyte.
- Technology Sack: Python, NLP, Deep Learning, Machine learning

Education

POST GRADUATION PROGRAM IN DATA SCIENCE ENGINEERING | MAR 19 - AUG 19 | GREAT LAKES INSTITUTE OF MANAGEMENT

- Machine Learning
- Data Analysis
- Deep Learning
- Feature Engineering
- Data Visualization
- Feature selection
- Natural Language Processing
- Python
- SQL

BACHELORE IN TECHNOLOGY IN EEE | 2014 - 2018 | SIR M VISVESVARAYA INSTITUTE OF TECHNOLOGY, BANGALORE

Skills & Abilities

- Language: Python (2yrs+), SQL (2yrs+)
- Technology: Machine Learning (2yrs+), Deep Learning (2yrs+), Natural Language Processing (2yrs+), Pipeline Designing (2yrs+).
- Code Versioning: GitHub (2yrs+), Bit-Bucket (2yrs+).
- API: Flask (2yrs+).
- Docker (beginner).
- Cloud Technology: AWS EC2 (2yrs+), S3 (2yrs+), SES (beginner) and SQS (beginner).
- Project Management: Jira (2yrs+).
- Visualization: Power BI (5 months), Tableau
- Applications: Postman(2yrs+), Py-Charm(2yrs+), MS Excel(2yrs+)
- Libraries: Panda, Numpy, PyTorch, TensorFlow, Sci-kit Learn, Transformers, NLTK, SpaCy, Keras, Seaborn, Flask.
- Database: MySQL(2yrs+), PostgreSQL(2yrs+)

Projects

1. Auto Analytics (Industrial Project) (AI).
    - Problem Statement: Develop a generic pipeline to generate all the reports for different clients.
    - Solution:
        i. Developed a highly customizable config file based pipeline which changes the code on real time according to the inputs it received.
        ii. Multi-sheet Excel file reporting and graphs are imbedded in the file as well.
    - Technology Stack: Python, Flask, Data Analysis, Power BI, AWS S3, AWS SE
2. Sygnif.ai - Document Processing Engine (Industrial Project) (AI).
    - Problem Statement: Extract information of KYC related documents.
    - Solution:
        i. Image classification model is used to classify the document is weather Passport, Pan card, Aadhaar card, Voter ID, Driving License etc.
        ii. Inception Net is used as base model and further fine-tuned according our requirement.
        iii. Object detection is used to classify image and QR-Code for further masking for customer's privacy.
        iv. OCR for text extraction.
        v. Rebuilding the text structure and classifying it on the basis of tags.
    - Technology Stack: Python, Deep Learning (DL), InceptionNet, PyTorch.
3. Symplif.ai - Claims Processing Engine (Industrial Project) (AI).
    - Problem Statement: Extract Information from insurance documents.
    - Solution:
        i. Apply functional model to classify the data types into information of interest i.e. Limit of insurance, Insurance Period, Deductibles, Listed Forms, Attached forms, States included/excluded.
    - Technology Stack: Python, Natural Language Processing (NLP), LSTM, PyTorch, Deep Learning (DL).

4. Insurance Learning Pipeline (Industrial Project) (AI).
    - Problem Statement: Re-training of the existing model periodically.
    - Solution: Developed an automated pipeline for training existing models into configurable periods and then deployed on the shallow-production environment where the performance of the new is compared with the existing model.
    - Technology Stack: Python, AWS S3, AWS SQS, AWS EC2
5. Resume Genie (Industrial Project) (AI).
    - Problem Statement: Extracting information of candidate and rank them.
    - Solutions:
        i. Trained a NER-based model to extract key information like Name, Email, Current Company, Last Company, Educational institutes, dates related to the organizations as well as institutes.
        ii. Check the TfIdf score of skill asked in JD with respect to the resume.
        iii. Build a model to classify resume as suitable candidate or not.
        iv. Using the information extracted via NER to rank them e.g. experience, institute, qualification etc
    - Technology Stack: Python, Natural Language Processing (NLP), Deep Learning (DL), Machine Learning (ML)
6. Facial Recognition based attendance system | Link (personal Project) (AI).
    - Problem Statement: Build a attendance system based on AI
    - Solution: Build a CNN based model which classifies face of students in 1-2 seconds with 96% accuracy and register their appearance in database along with timestamp and send notification via mail to their parent's/guardian's email at the end of the day if the student has been marked as absent.
    - Technology Stack: Python, Convolutional Neural Network (CNN), PostgreSQL, AWS SES
7. Electra based QnA deep learning Model | Link (personal Project) (AI).
    - Problem Statement: Build a model to extract question's answer from the provided context.
    - Solution: Build a flask based app which takes input a question and context (512 tokens) and extract answers from that context using fine-tuned Electra model.
    - Technology Stack: Python, Natural Language Processing (NLP), Deep Learning (DL), Transformers, Flask, AWS EC2.
8. Predicting power consumption of a house using Time Series Data | Link (personal Project) (AI).
    - Problem Statement: Predict the power consumption of a house based on its previous behavior.
    - Solution: Build a LSTM based model to forecast the power consumption of a housed based on its previous behavior.
    - Technology Stack: Python, LSTM, Time series analysis, Deep Learning (DL).
9. Facial Expression Classification | Link (personal Project) (AI).
    - Problem statement: Classify a facial expression.
    - Build a ResNet-50 for image classification.
    - Used haarcascade technique of face detection and drawing box around it.
    - Technology Stack: Python, Convolutional Neural Networks (CNN), Deep Learning (DL), Computer Vision (CV).
10. Patient Classification as Chronic patient | Link (personal Project) (AI).
    - Problem statement: Classify Patients with chronic kidney Disease.
    - Solution: Trained a Machine learning model i.e. KNeighbor Classifier to classify patients as diseased or not diseased with 97.91% accuracy and 97.23% recall and 100% precision.
    - Technology Stack: Python, Machine learning (ML), Sklearn (Sci-kit learn), Exploratory Data analysis, feature engineering, feature transformation.

11. Implementation of multiple SOTA model from scratch | Link (personal Project) (AI).
    - Implementing State of the art model from scratch (Architecture wise) for fun and deep understanding of the models.
    - Technology Stack: Python, Pytorch, Deep Learning (DL).
12. Kaggle work | Link (personal Project) (AI).
    - Solving various problem statements via applying machine learning on it.
    - Technology Stack: Python, Machine Learning (ML), Supervised ML, Unsupervised ML.
13. Convolutional Neural Networks | Link (Blog) (AI).
14. Deep Learning | Link (Blog) (AI).
15. Inblog | Link (Blog) (AI).
16. Visualisation of 2014 Elections | Link (Visualization) (Data).
    - Problem Statement: Develop an insightful dashboard for visualising the performance of all party in the 2014 election.
    - Technology Stack: Tableau.
17. Visualisation portfolio | Link (Visualization) (Data).
    - Tableau Public Portfolio of Data Visualization.
18. Hacker Rank | Link (Competitive Coding).
19. Code Blogs | Link (Blog) (Competitive Coding).

Certifications

1. Specialization Course on Deep Learning offered by Deeplearin.ai and Coursera | Link
    a. Improving Deep Neural Networks
    b. Structuring Machine Learning Projects
    c. Convolutional Neural Networks
    d. Neural Networks and Deep Learning
    e. Sequence Models
2. Specialization Course on Natural Language Processing offered by Deeplearning.ai and Coursera | Link
    a. Classification and Vector Spaces
    b. Probabilistic Models
    c. Sequence Models
    d. Attention Models
3. Python Data Structures offered by University of Michigan and Coursera | Link

Achievements, Activities and Interests

1. Employee of the month (August 2021) (Recruitment Smart Technology Pvt. Ltd.).
2. Employee of the month (Feb 2021, Dec 2020 and Nov 2020) (Sumyag Datascience Pvt Ltd)
3. Running, Yoga and Playing Badminton.