Shiwangi Pasari

shiwangipasari@gmail.com

LinkedIn: shiwangi-pasari | +91-9123073988

PROFILE SUMMARY

- **5 years 7 months of experience in software development with major focus in Machine Learning, Python, Backend development, Microservices, RestAPIs, Natural Language Processing, Messaging Queues, Flask, MongodB and related technologies.**
- **Experience in developing NLP based models for text extraction, classification.**
- **Basic understanding of AWS and used its services lke Lambda, DynamoDB, SNS, S3.**
- **POC on Flood Alarm System using IOT and AWS services like Lambda, DynamoDB, SNS and a bit of other services.**
- **Expertise in writing python and pandas code, performance improvement of python code and code structuring/designing.**
- **Excellent time management skills with ability to perform under pressure & meet deadlines.**
- **Ability to learn and adapt quickly to emerging new technologies and paradigms. Learnt many technologies on job as per the project requirement.**

EDUCATION

Indian Institute of Technology, Hyderabad   Hyderabad, Telangana
Master of Technology in Data Science   *Aug 2021 - Present*

Academy of Technology (MAKAUT, Formerly known as WBUT)   Kolkata, West Bengal
Bachelor of Technology in Computer Science and Engineering; GPA: 9.16/10.0   *Jul 2013 - May 2017*

PROFESSIONAL EXPERIENCE

Software Developer II - Machine Learning   July 2022 - Present
Google   *Bangalore, India*

Development Team Lead   October 2017 - July 2022
Persistent Systems   *Bangalore, India*

SKILLS

- **Languages**: Python, C, SQL, Java
- **Web Technologies**: Flask, Django-Rest
- **Operating Systems**: Windows, Linux
- **Tools**: Git, SVN, JIRA, Eclipse IDE, PyCharm, Jupyter
- **Libraries**: Scikit-Learn, Numpy, Pandas, Geopy, OpenCV, NLTK, spaCy, Plotly, Seaborn, Google vision, Matplotlib, Keras, word2vec, XGBoost, BeautifulSoup, Flask, MongoDB, stomp.py, asyncio, polyglot, GATE(General Architechture for Text Engineering), multiprocessing, Flask-caching, grequests, gunicorn, image-magick, Apache-Nifi, Apache-Airflow.

PROJECTS

- **Google One**:
  Google One is the subscription service developed by Google that offers extra cloud storage and other benefits to its users. Its a paid service that provides an expansinon from the free 15GB storage Google account to a maximum of 30TB storage. It is shared across multiple Google products like Google Photos, Gmail, Drive, etc.

  Roles and Responsibilities

    - Worked on Unification of Database for the pipeline.
    - Worked on setting up ITS server for functional test of the component.
    - Worked on optimization and re-structuring of the code.
    - Was a part of team for Benefits Personalization and experiment launch.
    - Working on ML based notification service.

  Technologies used: Java, Python, ITS, Tfx, Spanner DB, Cider-v, Plx

- **Vitrina Platform Engineering**:
  Vitirina.ai is envisioned to be a B2B SaaS platform that provides a suite of applications. The application enables users to search for information for content and companies using curated information from select news articles and industry catalogues. The application also provides a way to visualize the relationships within the various entities as a connected knowledge graph.

  Roles and Responsibilities

- Designed and developed different Python modules.
- Developed relevant rest APIs for the services.
- Worked on derivation and extraction of important entities and attributes.
- Analysed various possibilities on relation classification challenges and designed the possible approach.
- Working on various relation classification techniques and other optimization tasks.

Technologies used: Python, Pandas, spaCy, Scikit-Learn, Numpy, Plotly, Seaborn, Matplotlib, Keras, word2vec, BeautifulSoup, selectolax, gensim, nltk, fuzzywuzzy

- **CISCO Project**:
CISCO Case Deflection is a data science solution to reduce active issues raised by their customers. It uses NLP techniques to provide on-the-fly appropriate documents relevant to the query, so that customer issue is solved before registering it in the system.

  Roles and Responsibilities

  - Worked on different subprojects to develop and generate different analytics report.
  - Worked on optimizing the performance for the python module.
  - Worked on generating relevancy score using ndcg_score.
  - Used various similarity techniques to generate string and document similarity scores.
  - Working on important keyword extraction similar to text summarization of domain specific problem using various NLP techniques.

  Technologies used: Python, Pandas, Scikit-Learn, Numpy, Plotly, Seaborn, Matplotlib, Keras, word2vec, BeautifulSoup, selectolax, gensim, nltk

- **ADMS - Automated Digitization & Management System (Product Development)**:
ADMS is a solution to handle the workflow of digitizing document, annotating the entities of interest and providing a search over the documents. ADMS aims at building a reusable, scalable, extensible and highly accurate digitization and search system.

  Roles and Responsibilities

  - Designed the workflow of orchestration engine as per requirement.
  - Implemented ocr generation engine using google-vision. Also, added support for multi-column ocr text parsing in OCR.
  - Created NER models using spaCy for Hindi and English documents.
  - Independently created microservices, rest end-points for different components and object-pooling of a component.
  - Independently implemented various features in NER, Ebook and OCR.
  - Extensive use of MongoDB to perform different CRUD and aggregate operations. Implemented versioning of documents in MongodB.
  - Handled encryption of sensitive data, caching at server level.
  - Independently added support for the components in Airflow.

  Technologies used: Python, Pandas, Flask, stomp.py, pymongo, ActiveMQ, MongoDB, spaCy, Google-vision, levenshtein, fuzzy-wuzzy, Image-Magick, Gunicorn, Waitress, Apache Airflow, GATE, java, hiberante, spring

- **Text.AI - Sanctions Entity Feedback Learning (Platform Development)**:
This use case was a typical Feedback learning project which boosted accuracy and coverage of Sanctioned Entity use case. Sanctions Entity Feedback Learning had three feedback features namely - Correction, Additions and Context. It is also capable to learn and improve on its own by accepting the feedback from the business user thus improving coverage and accuracy.This solution 0.4 million user data each month efficiently.

  Roles and Responsibilities

  - Independently written code for corrections, context and additions models.
  - Designed and developed the code base and model structures under strict deadlines.
  - Applied rollback, model backup and cold deletes functionality.
  - Written the feedback learning implementation for ensembler (main component).
  - Optimized performance time by 300%(approx).
  - Held responsibility to make code builds with complete integrity and commitment.

Technologies used: python, pandas, numpy, levenshtein, fuzzywuzzy, NLTK, scikit-learn, datetime, num2words, spaCy, keras, jupyter, GATE, java, hiberante, spring

- **Text.AI - Collections (Platform Development)**:

  This use case was similar to aforementioned use case. It included extraction of 67 entities, extraction and management of data from check-marked document and document classification

  Roles and Responsibilities

    - Trained spaCy model for Entity extraction of 15+ entities using client's 50,000 text data.
    - Written GATE rules and mentored team members on text extraction using GATE.
    - Analyzed client's data for model building and useful pattern making.
    - Written many complex business logic for tenor and extraction of document attached entity (important business entity).
    - Improved Document classification accuracy from 86% to 94%.
    - Performed Python back-end code optimization to reduce execution time by 46%
    - Held responsibility to make code builds with complete integrity and commitment.
    - **Production Accuracy: 82%, Coverage: 90%**

  Technologies used: python, pandas, numpy, levenshtein, NLTK, scikit-learn, datetime, num2words, spaCy, keras, jupyter, GATE, java, hiberante, spring

- **Text.AI - Sanctions Entity (Platform Development)**:

  Used Machine Learning driven Natural Language Processing platform (Text.AI) for Named Entity Recognition, Classification and Intelligent decision making. The platform is capable of extracting 5 entities namely - PERSON, LOCATION, ORGANIZATION, VESSEL and PORT. The model was intelligent enough to denote country code and information like sanctioned country, restricted country and restricted organization.

  Roles and Responsibilities

    - Implemented various modules on python using pandas library for cleaning and analyzing data.
    - Applied robust and object oriented approach while writing code.
    - Used extensive regex and beautifulsoup libraries for web extraction to make dictionaries.
    - Worked on XML parsing in Java.
    - **Production Accuracy: 90%, Coverage: 95%**

  Technologies used: python, pandas, levenshtein, java, GATE, hibernate, spring

- **IGRS - (POC)**:

  IGRS is entity extraction and classification project. The main challenge of extracting information is that the text files are present in English, Hindi and Bilingual languages. Here, We extract property specific details using Machine Learning and Natural Language Processing.

  Roles and Responsibilities

    - Created spacy model on client's data for English and Hindi documents.
    - Implemented rules for text extraction from GATE.

  Technologies used: python, pandas, numpy, levenshtein, NLTK, scikit-learn, datetime, num2words, spaCy, keras, jupyter, GATE, java, hiberante, spring

- **Thermofisher PSD (Product Development)**:

  Web based application to ease store of PSD information into the database and help provide different analytical report on their data.

  Roles and Responsibilities

    - Lead the backend development of rest-apis and analytical report.
    - Worked on complex queries to provide proper report on the existing data.
    - Worked in coordination with different teams to deliver before deadline.

  Technologies used: Python, Pandas, Django-rest, Gunicorn, SQL

ADDITIONAL EXPERIENCE & ACHIEVEMENTS

- **Participated and won as a team in Semicolons-2019** (Hackathon organized by Persistent Systems), The competition was held worldwide with 47 teams participating in the event.
- Developed and delivered the back-end code of **sales.AI(semicolons project)** within 24 hours
- Identified as Top Talent in the organization for consecutive 3 years