Name: Vaibhav Hiwase
Designation: Senior Consultant in Data Science
Qualification: Master of Technology in Computer Science

Background

- **4** years of experience across various verticals
- Research and develop algorithms/systems for given problems independently
- Working in a highly collaborative environment with teams to deliver systems from prototyping to production level
- Expert in Natural Language Processing, Azure Cognitive Services and Azure custom vision
- Solving real-world scenarios for user commands and requests and building scalable systems that solve their problems
- Team management
- Experienced in performance monitoring, load balancing, CI/CD pipeline
- Expert in Containerization and creation of production ready multi-stage build of Dockerfile
- Good hands-on experience in data analysis and manipulation tools like pandas, and python regular expressions
- Expertise in Azure container app, Azure, Azure API Management, Azure Database for MySQL, Azure Cache for Redis and AKS
- Experienced in Providing technical assistance to improve system performance, capacity reliability, and scalability.

Skills

- Expertise in Kusto Query Language
- Expertise in Optical Character Recognition
- Knowledge of Azure Configurations and Azure Networking, Azure data explorer
- Security, Scalability and Performance monitoring
- Azure Monitor
- Cost monitoring
- Automation tools
- Key Vault
- Storage Account
- Python, Flask, Pandas and Regular Expressio
- Redis, Redis Queue and Callback URLs
- Architecture Diagram
- Data Flow Diagram
- Problem Solving
- Critical Thinking

Published Research Papers

- Dimensionality Reduction for Improving the Performance of Risk Calculation Using Machine Learning Algorithms
  - Publication: HELIX, 2018
- Review On Application of Data Mining in Life Insurance
  - Publication: International Journal of Engineering & Technology, 2018

Relevant professional experience

- **Key Information Extraction**
  - Handled complex document layout by training a deep learning model for key information extraction from documents
  - Used improved graph learning-convolutional networks for receipts and KYC documents with above 94% accuracy.
- **Document Digitalization:**
  - Role: Research, Engineering, Development, Architecture designing, Infrastructure setup, Network connectivity, Application Monitoring, etc.
  - Created a proprietary solution for payslip forgery detection using metadata and pixel-level analysis and detect forgery from additional layer creation as well as incremental updates in editable PDFs.
  - Responsible for Setup Azure landscape creating Infrastructure which includes Virtual Network, Subnet, API Management, Virtual Machine, Key Vault, Storage Account, Application Insight, Container registry and Kubernetes Service.
  - Removed public access of AKS load balancer (which is not secured) and attached it to the secured gateway endpoint with SSL binding integrated with Azure API Management service instance
  - Applied policy to manipulate, request and add useful features such as IP filters which can block and allow traffic from allowed IP addresses

- Added some security headers to validate all API calls in inbound traffic. These security tokens are safely stored in azure key vault, so that we can change them without any downtime
- Handled Long-running task which was part of application's workflow in the background, outside the normal flow using Redis Queue
- Distributed load across all available worker-pods of AKS in asynchronous REST API calls to handle long running task which include processing of more than one PDF in single endpoint
- End to End Support for their infrastructure Setup.
- Monitoring of application using application insights and Kusto Query Language for Azure Logs
- Network hardening
- Developed the solution for document similarity using TF-IDF and Optical Character Recognition

- **Custom Vision Dashboard Training:**
  - Role: Delivering system from prototyping to production, Infrastructure setup, Network connectivity, CI/CD, Team management, Application Monitoring, etc.
  - Created the architecture of solution using Azure Container app service, GitHub Action, Azure Container Registry, Azure Database for MySQL, Azure Cache for Redis and Azure Monitoring
  - Created data flow for Custom Vision model training pipeline
  - Collaborated with team to deliver systems from prototyping to production level
  - Created production ready multi-stage build for Dockerfile which can communicate backend built in React app and frontend built in Flask app in a single container deployment instead of two which reduces cost by half
  - Followed best practices in software development

- **WEAV AI**
  - WEAV AI is a no-code open-source platform for data scientists
  - Implemented the APIs for data ingestion services in WEAV Core libraries using Airbyte connectors

- **Smart Content Extraction:**
  - Proposed a state-of-the-art approach for paragraph extraction and table extraction using a clustering algorithm
  - Performance optimization using multi-core computing
  - Lead and established the solution architecture, work breakdown structure (WBS) with a team of 10 responsible for the development and value-driving implementations of Smart Content Extraction
  - Created and deliver an end-to-end project using Flask API, Docker, GIT, and Azure Cloud
  - Implemented a Redis Queue to handle asynchronous services in the Flask app and create tasks
  - Containerized Flask and Redis with Docker and separate a worker process for long-running tasks in the background
  - Created setup to monitor queues, jobs, and workers using RQ Dashboard
  - Ensured the integration of service delivery on the customer's requirements

- **Recommendation System**
  - Built backend for keyword identification, clustering, recommender for media analytics

- **Computer Vision**
  - Successfully developed and trained LeNet Deep learning model to perform pixel-level pathogen classification of Fungus, Yeast, and Bacteria on the histopathological stain of the human nail tissues
  - Handled > 5 TB image dataset in this project developed from histopathological techniques on the stain of Grocott methenamine silver (GMS), Periodic acid-Schiff (PAS), and Hematoxylin and eosin stain (H&E).
  - Performed segmentation on slides (images) to create binary masks using OpenCV

- **Automatic Speech Recognition**
  - Trained and implemented Baidu's Deep Speech 2 (Tensorflow) in Python3 with Common Voice dataset to assist the creation of speech recognition solutions for the healthcare industry

- **Optical Character Recognition**
  - Digitizing contracts
  - Developed predictive models to extract key features like obligations and clauses in the contract-based documents
  - Worked on the project module to extract meaningful structure from PDF documents along with page header-footer removal

- **Face Recognition**
  - Recognizing faces in images using CNN architecture and analyzed a histogram of oriented gradients (HOG) for face recognition
  - Compared face encodings using similarity algorithms
  - Performed pre-processing operations on identity cards image for OCR and performed attribute extraction using NER
  - Parsed expected information from hOCR format. EDA using Pandas