Vaibhav Hiwase

Senior Consultant with 3+ years of experience in machine learning and python software development. A personable, adaptable, well-rounded professional who specializes in NLP, DevOps and Azure cloud solutions.

hiwase.vaibhav@gmail.com
+91 8999382164
Nagpur, Maharashtra, India
LinkedIn
Website

Work experience

09/2021 - Present — Nagpur, India
Senior Consultant in Data Science
Celebal Technologies

08/2018 - 08/2021 — Nagpur, India
Machine Learning Engineer
Konverge AI Private Limited

Open Source

- Author of matrix-rotation PYPI package.
- Author of path-traveler library.

Employment verification from Payslip

- Delivered a custom software development tool on time and help to build proprietary solution for payslip forgery detection using metadata and pixel-level analysis and detect forgery happened from additional layer as well as incremental updates in editable and scanned PDFs.
- Automate manual, time-consuming, and broken process prone to human error, reducing delays in customer service and increasing number of frauds detection by verifying user's employment data in less than 60 seconds.

(Python3, Docker, Azure Kubernetes Service, Azure DevOps, Azure Custom Vision, and Azure Cognitive services)

News Topic Identification

- Data gathering from web scrapping
- Training a BERT model using semi-supervised topic modeling machine learning algorithm and categorized then into 109 different categories to automate manual assignments.
- Building a news article recommendation system from unsupervised training on BERT Topic model

(Python3, Docker, Azure Kubernetes Service, Azure DevOps)

Smart Content Extraction

- Developed python APIs for WEAV.AI data Ingestion services.
- Project Lead in Digitizing contracts.

Education

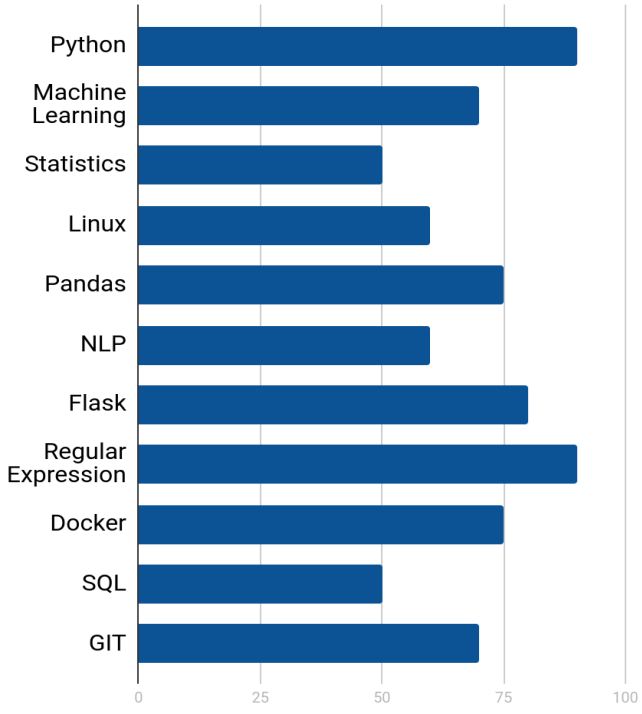
2016 - 2018 — Nagpur, India
Master of Technology (MTech)
RCOEM
Computer Science and Engineering

08/2012 - 06/2016 — Nagpur, India
Bachelor of Engineering (B.E)
SVPCET
Information Technology

Strengths

Learning Agility | Lateral Thinking | Critical Thinking | Problem Solving

Skills and Competencies



- Proposed a state-of-the-art approach for paragraph extraction and table extraction using a clustering algorithm.
- Lead and established the solution architecture, work breakdown structure.
- Created and delivered an end-to-end project using Flask API, Docker, GIT, and Azure Cloud and optimize performance.
- Implemented a Redis Queue to handle asynchronous services in the Flask app and create tasks, RQ Dashboard (to monitor queues, jobs, and workers)
- Containerized Flask and Redis with Docker and separate a worker process for long-running tasks in the background.
- Developing predictive models to extract key features like obligations and clauses from contracts using SVM model.
- Attribute extraction using NER and validation approach from respective clauses.
- Worked on the project module to extract meaningful structure from PDF documents along with page header-footer removal.

(Python3, scikit-learn, SVM, TF-IDF, Pandas, NLTK, Flair, Spacy, BeautifulSoup, Decorator, Docker, Redis, Flask, Azure VM, Azure READ API, PDFtk, OpenCV, Imagemagick, Camelot, Tabula)

eKYC Project

- Recognizing faces in video frames.
- Analyzing a histogram of oriented gradients (HOG Model)
- Comparing face encodings using similarity algorithms.
- Perform preprocessing operations on identity cards image for OCR.
- Parsed expected information from hOCR format. EDA using Pandas.
- Developed the backend of the application using the flask Framework.

(Python3, Object detection, Semantic. Segmentation, Tensorflow, Keras, OpenCV, Tesseract 4, PDFtk, Imagemagick)

Medical Image Segmentation and Detection of Pathogens

- Performed segmentation on slides (images) to create binary masks using OpenCV.

(Python3, Flask, Dlib, Tesseract, OpenCV, Pandas, Numpy, pickle, regex, KNN)

Video Analytics Platform

- Built backend for keyword identification, clustering, recommender for a media analytics organization.

(Python3, Shell, sklearn, Pandas)

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

More Details

QGIS Platform

- Built system for correlating text with geographical locations from produced scanned govt documents.

(PyQGIS API, Pandas)

Speech-to-Text Platform

- Worked on Automatic Speech Recognition Engine (ASRE) for healthcare application.
- Implemented Baidu's pre-trained DeepSpeech2 model in Python3.
- Trained Mozilla's DeepSpeech Model in Kaldi using a common voice dataset.

(Keras (TensorFlow), DeepSpeech2)

08/2016 - 06/2018 Nagpur, India

Master's Degree (Engineering)

RCOEM

Computer Science and Engineering

- Performed research and analysis in the life insurance sector using machine learning and data statistics.
  Review on Application of Data Mining in Life Insurance
- Assisted in research for classification of ~2000 research paper titles describing the biological function of Extracellular Vesicles (CV)
- Explored the state of art and scalable solution of movie rating recommendation system using stacked autoencoders (DNN)