|  |  |  |
| --- | --- | --- |
| **LALIT ASHOK PATIL**  lalit\_patil2@persistent.com |  | |
| **Profile Summary**   * Experience in Python, Machine Learning, Deep Neural Network, Image Processing, Microsoft Azure and Generative AI * 7+ years of experience in Machine Learning model framework development, developed and deployed end to end Machine Learning model pipelines * Extensively worked on deep Learning and machine learning use cases and had delivered model built on tensorflow, keras, pycaret, mlflow, sklearn etc * Conceptualized, developed of Linear Regression, Random Forest, Support Vector Machine, Decision Tree, Cluster, Cosine Similarity Recommendation System ML models and CNNs model in deep neural netwok * Efficient in developing data insight and visualization from raw data using SQL, Python pyspark and Excel * Experience in NLP, LLM, Generative AI, Prompt Engineering, Retrieval Augmented Generation and Code Conversion * Experience in pulling data from ADLS, big data and data refinement * Acquired various certifications like Machine Learning, Python, Deep Neural Network, Data Analytics and Databricks Certifications * Experience in research and development and paper publications in journals   **Work Experience**  **Protocol Digitalization, Jan 2024 – till date**  **Role**: Data Scientist  Conducting clinical trials for drug development is an expensive but critical step to get patients their needed treatments. On average, each trial ends up with over 2 amendments during the trial, which can result in $0.2 – 1+ million of additional costs. Improving the design process from the start can build the foundation to prevent additional costs. This process can be tedious, complex, and requires data from a wide range of sources.  The solution is designed and implanted with AI, specifically NLP, can unearth the required data effectively at scale to help support, plan, and design clinical trials. It consists of a blend of tools, including linguistic rules, ontologies, machine learning (ML), and large language models (LLMs), which are used to find insights from text. It can be applied to multiple sources including ClincalTrials.gov, protocols (public or proprietary), publications, social media, and more.  **Responsibilities**:   * Requirement elicitation & analysis * Defining the schema and implementation for data refinement * Implement the Generative AI pipelint to extract the healthcare data.   **Environment**: Windows 10, ubuntu  **Worked Tools/Technologies:** OenAI Chatgpt 3.5 and 4, ML, NLP, LLM, RAG, LangChain, Python, Pandas, PostgreSQL, Gitlab, Jira, ALM, VS Code  **Audience Measurement, Aug 2023 – Dec 2023**  **Role**: Data Engineer  Audience measument is post panel exchange data refinenement operations for media data. To pull the big data from different sources. Data ingestion is done through ADF and refinement with databricks. The Hive database is use for storing and processing big data.  Airflow is used to orchestrate the overall project flow.  **Responsibilities**:   * Requirement elicitation & analysis * Defining the schema and implementation for data refinement * Implement the ADF pipeline to pull the meta data and airflow for orchestration   **Environment**: Windows 10, ubuntu  **Worked Tools/Technologies:** Azure Databricks, Azure Data Factory, Pyspark, Azure Data Storage, Python, Airflow, PyCharm  **Rapid Data Extraction Using GenAI, Aug 2023 – Oct 2023**  **Role**: Data Scientist  The solution takes the variety of documents in different formats like PDF, image etc. The information(structured/unstructured) from these documents is parsed using Azure AI Services/open source libraries. The solution is having additional image processing layer to improve the quality of distorted/noisy images and OCR is performed to extract information from the scanned images/documents. The solution is having entity extraction engine using GenAI. The content parsed from the uploaded documents is passed to the engine , which will Identify the Document Language Identify the Document type Extract the required headers and line items from the document. These extracted entities are displayed to users, and they will be able to regenerate the response if engine misses any fields. The goal of solution is to assist user with efficient document processing which increases their productivity and efficiency.  **Worked Tools/Technologies:** Databricks, Azure OpenAI, ChatGPT3.5, azure formrecognizer, streamlit, python, PyCharm  **Informatica to StreamSets migration Using GenAI , July 2023 – Sept 2023**  **Role**: Data Scientist  Informatica to StreamSets migration to speed up the work with accuracy and efficiency. The iterative approach for prompt engineering is applied to optimize the efforts.  **Worked Tools/Technologies:** Databricks, Azure OpenAI, ChatGPT3.5, python, Pycharm, Azure Databricks  **Service Modernization for Grocery, Dec 2021 – June 2023**  **Role**: Data Scientist  Service Modernization project revolutionizes how an enterprise provides service and support to its customers by leveraging advancements in computing and connectivity, proliferation of connected things, and maturity of technologies such as artificial intelligence that will drive an increasing reliance on machines.  **Responsibilities**:   * Requirement elicitation & analysis * Data gathering and dataset preparation * Planning, Designing and implement machine learning models in accordance with customer needs * Investigate, test, fine tuning and put into practice appropriate ML algorithms * Implementation of production model, testing and deployment * Customer communication and team support   **Environment**: Windows 10, ubuntu  **Worked Tools/Technologies:** Python, Machine Learning, Azure Data Storage, Azure Databricks, Azure Dev, Pyspark, Cassandra, SQL, Pandas, Pyspark, pycaret, sklearn, mlflow, Feature Storage, Airflow, VS Code  **Bill of Lading Management, Oct 2020 - Dec 2021**  **Role**: Sr. Python Developer  The solution intends to provide intelligent document processing to extract data from bill of lading, integrate with core systems, and trigger auto-mails to concerned parties. The solution is incorporated with auto document splitter and Bill of Landing Identification using Deep Neural Network.  **Responsibilities**:   * Requirement elicitation & analysis * Technical Design * Data gathering and dataset preparation * Design and implementation of training model * Measure and improve results of training models * Implementation of production model, testing and deployment * Customer communication & team support   **Environment**: Windows 10  **Worked Tools/Technologies:** PyCharm, Miniconda ,Python, Deep Neural Network, Tensorflow, Keras, Pillow, Pyodbc, CV2  **TruCap+ Intelligent Document Processing - June 2019 – Sept 2020**  **Role:** Sr. Python Developer  The solution intends to provide Intelligent Document Processing or IDP (Intelligent Document Processing) is a technology solution for extracting data from unstructured paper forms and assets and converting it into a structured form that can be used for further processing towards end-to-end automation. The solution is incorporated with auto document identification and language identification using Machine Learning and Deep Neural Network.  **Responsibilities**:   * Requirement elicitation & analysis * Technical Design * Data gathering and dataset preparation * Design and implementation of training model * Measure and improve results of training model * Implementation of production model, testing and deployment * Customer communication & team support   **Environment**: Windows 10  **Worked Tools/Technologies:** Spyder, Miniconda, Python, Deep Neural Network, Machine Learning, Tensorflow, Keras, Pillow, Pyodbc, CV2, RabbitMQ  **Research and Development - Feb 2011 – June 2019**  **Role:** Assistant Professor  The work intends to do Research and Development to raise funds from various government institutes and organizations. Along with research related activities conducing training, laboratories, and lectures. Project guide, implementation and publications with UG and PG students. Learning various tools and technologies and included into the laboratory and syllabus.  **Responsibilities**:   * Research and Development * Research Publication * Training, conducting laboratory and lectures * Book publication * Project guide and implementation   **Environment**: Windows, Ubuntu  **Worked Tools/Technologies:** Spyder, Miniconda, Python, Deep Neural Network, Machine Learning, Tensorflow, Keras, Pillow, Pyodbc, CV2, Information Retrieval, Internet of Things, Artificial Intelligence, Pattern Recognition and Computer Networks  **Education**  Master of Engineering (Computer Engineering) – K. K. Wagh IEE&R, Nashik - 2012  Bachelor of Engineering (Computer Engineering) – AISSMS, COE, PUNE - 2010  **Certifications**   * NPTEL online certification of “Wireless Ad Hoc and Sensor Networks course”, 2017 (Completed with “Elite” category with 77 marks out of 100) * Online certification of “python”, Big Data University, 2017. * NPTEL online certification of “Introduction to Machine Learning”, 2017. * NPTEL online certification of “Introduction to Internet of Things”, 2018. (Completed with Top 1% all India Rank-87 marks out of 100) * NPTEL online certification of “Introduction to R Software”, 2018. (Completed with Gold Category-94 marks out of 100) * Online certification of “Introduction to R programming”, Big Data University, 2018. * Online certification of “Machine Learning with R”, Big Data University, 2018. * NPTEL online certification of “Data Analytics”, 2020 * Machine Learning Certification – Analytics Vidhya – 2021 * Fundamentals of the databricks lakehouse platform accreditation - Databricks Certification - 2022 * Deployment of machine learning models certification - Udemy -2022 * Databricks Certified Machine Learning Associate - Databricks Certification – 2023 * Databricks Certified Machine Learning Professional - Databricks Certification – 2023 * Langchain-develop LLM powered applications with langchain - Udemy - 2023 * GenAI for Practitioners – Persistent University – 2023 * Generative AI with Large Language Models- Coursera - 2023   **Awards and achievements**   * High-Five Individual Awards for contribution towards GenAI Initiatives - Issued by Persistent Systems Limited, Sept 2023 * Top Talent - Issued by Persistent Systems Limited, May 2023 * Bravo Award - Issued by Persistent System Limited, Mar 2023 * Semicolon winner - Persistent Systems Limited, March 2022 * Leadership in business excellence – BOL Team, 2021 * GATE Qualified - Issued by IIT Bombay, May 2010   **Publications**   * Lalit A. Patil, Prof. S. M. Kamalapur, “Web page Clustering Using Latent Semantic Analysis”, International Journal of Computer Application -2012. * Lalit A. Patil, Prof. S. M. Kamalapur, Dhananjay M. Kanade, “Improving Web page Clustering Using Latent Semantic Analysis”, International Journal of Computer Application -2012 * Lalit A. Patil, Prof. S. M. Kamalapur,”Improving Web page Clustering Using Multiview Learning”, in International Journal of Emerging Trends & Technology in Computer Science, (ISSN 2278-6856), Volume 2, Issue 2, March - April 2013. * Prof. D. M. Kanade, Lalit A. Patil, “Internet of Things Security: Challenges and Opportunities”, in IJRAR, ISSN: 2321-9653, IC Value: 45.98, SJ Impact Factor: 6.887, Volume 6, Issue V, May 2018. * Lalit A. Patil, Prof. D. M. Kanade, “Internet of Things Standardization: Challenges and Opportunities”, in IJRAR, ISSN 2349-5138, SJ Impact Factor: 5.75, Volume 6, Issue I, Jan 2019. * Shaswat Babhulgaonka1, Jayesh Suryavanshi, Pritam Bendkule, L. A. Patil, “Prediction and Diagnosis of Cardiovascular Diseases Using Machine Learning: A Review”, ISSN 2349-6002, Volume 5, Issue XII, May 2019. * Lalit Patil, Satish Wagh, “Computer Networks”, Vishwakarma Publication, Pune, January 2015, ISBN 9788383572618. | | **Key Skills and Knowledge**  Machine Learning  Deep Neural Network  Natural Language Processing  Generative AI  LLM  Python  Pyspark  Pandas  Mlflow  Azure Data Storage  Azure Databricks  Azure OpenAI  **Other Skills**  Tensorflow  Keras  Pytorch  LangChain  Prompt engineering  Vectore Database  Pycaret  Sklearn  Numpy  Data Visualization  Airflow  Hive  Cassandra  PostgreSQL  Databricks feature storage  Azure Data Factory  Azure Dev  RabbitMQ  Product Development  **Tools**  PyCharm IDE  VS Code  Spyder IDE  Jupyter Notebook  Databricks Notebook  Github  Gitlab  Jira  ALM |

Declaration: I hereby declare that the above provided information is correct to the best of my knowledge and belief.

Lalit Patil