

Parag Patil

Lead Data Scientist and data science certified professional with 6 years of experience executing data driven solutions to increase accuracy, efficiency and utility of internal data processing, Experienced at creating data regression, classification and NLP models using predictive data modeling and analyzing data using different tools to deliver insights and provide solutions to different complex business problems.

E1102, Tanish Orchid, Charoli,
Pune-412105

+91 9975555279

parag.patil5555@gmail.com

<https://www.linkedin.com/in/parag-patil-628372105>

EXPERIENCE

Harbinger System, Pune— Tech Specialist- AI/ML

June 2021- Present

- Working on designing and development of AI/ML framework for Healthcare, EdTech and HRTech domain.
- Working closely with VP's, Business Head to define scope of AI/ML techniques as per industry need.
- Working on building AI/ML accelerator's using NLP, Computer Vision and advanced data analytics techniques.
- Working with multiple clients to leverage AI capabilities in their existing product and services.
- Doing research on new techniques like NLP, GAN's, MLOps and AutoML

AI/ML Projects

- Slimthru Healthcare- Generating medial notes on Video/Audio healthcare transcription data with medical codes recommendation using NLP
- Candidate Recommendation- Candidate recommendation based on job description using finetune BERT language models
- Content Curation- Advance context based search engine using NLP and Elasticsearch.
- Contactless Attendance System- Attendance system using computer vision.
- Behavior Assessment Scoring- NLP based assessment scoring system.

Absolutdata, Pune— Consultant - Data Science

March 2021- June 2021

- Worked on automating existing AI pipelines using Apache Airflow.
- Improving existing NLP models using SOTA models like GPT3, RoBERTA

Exela Technology, Pune— Data Scientist

September 2019- March 2021

- Worked on NER extraction from legal and contract documents using custom train models like Spacy, Flair.
- Build a deep learning NLP model that can work as text summarization on legal documents using the T5 transformer model.

TECHNICAL SKILLS

Machine Learning (Classification, Regression, Feature Engineering)

- **Algorithm**
 - Random forest, XGBoost.
 - Linear regression, Logistic Regression
 - k-means

AI (Deep Learning and Neural Network)

- **Framework**
 - Keras
 - Tensorflow
 - Pytorch
 - Apache Airflow
- **Architecture and model**
 - Spacy
 - RASA
 - BERT
 - T5
 - CNN

Programming Language

- **Python**
 - Scikit-learn
 - Pandas, Numpy, NLTK.
 - Pymupdf, PIL, OpenCV,ocrmypdf
 - Selenium
 - Flask, FastAPI

- Converted previously built NLP model (sec-vec) that extracts clauses (sentence) from text document into more accurate text generation model (sec-sec).
- Replace manual entity extraction and redaction on PDF documents with automated entity extraction and redaction model using combination of pre-trained and costume train model.
- Worked on document classification model using architecture like BERT and T5 (Text-To-Text Transfer Transformer)
- Use scrapping on public government websites to gather information and documents.
- Developed generic captcha breaker using CNN and OpenCV for scraping data from websites.
- Created Web services using Flask, Gunicorn and Nginx.
- Successfully deployed multiple ML projects on QA and UAT environments using Docker.

AI/ML Projects:

- Jet Legal: NER extraction, Entity Redaction and text summarization.
- Contract Management Tool: Document classification, clauses extraction and language generation.
- Capta Detection Using CNN and OpenCV.
- PDF Redaction Using NER.
- RASA chatbot for Assistance.

Tech Mahindra, Pune— Associate Data Scientist

May 2017 - September 2019

- Process large-scale data using python (pandas, numpy) and other database management tools.
- Analyze data to find patterns or trends using Python's different libraries like pandas.
- Developing statistical models that relate to improvement in overall process, leads to 20% increase in process speed.
- Helping in preparation of data for conducting exploratory and predictive analytics using libraries like pandas, matplotlib and algorithms (CART, regression, SVM etc.)
- Helping in extracting important information from unstructured data like email using NLP with the help of regex and nltk libraries. With this it saves 10% overall time for email reading without loss of important information.

AI/ML Projects:

- ML classification model for server will patch or not patch
- NER extraction using Spacy.
- Classification of email as per importance
- Predicting server failure using survival analysis

Tech Mahindra, Pune— Data Analyst

September 2016 - May 2017

- Assist in the creation of reports and online visualization using Power-Bi and Tableau.
- Gather and process data using MySql and Python.
- Helping in automation of reports, these were previously prepared and sent manually.
- Migrating existing excel reports into Tableau dashboard.
- Presenting analytics reports to clients.

Tableau, Power Bi

Docker, Nginx

AWS- S3, EC2 and Sagemaker, comprehend

Azure- Blob, Event Hub, VM

SQL, MongoDB, Elasticsearch

AWARDS

2 Superstar Award -Harbinger

Best Performing Team Award -Harbinger

ACE Awards 2019 -Tech Mahindra
(Best performer of year)

EDUCATION

MS. Data Science
LJMU University, UK
Dec 2021 - Present

B.E. Computer Science
DPCOE, Pune University
May 2011 - May 2016

CERTIFICATION

Data Science Pro-Degree Genpact 2018
Imarticus Institute- Pune

NLP - Natural Language Processing using Python
Udemy-2019

Deep Learning and Neural Networks
Udemy-2019