

Mahesh kumar katakam

Associate Data Scientist

Data Scientist familiar with gathering, cleaning and organizing data for use by technical and non-technical personnel. Advanced understanding of statistical, algebraic and other analytical techniques. Highly organized, motivated and diligent with significant background in Machine Learning and NLP.

✉ maheshlucky9932@gmail.com

📞 8072180530

📍 Nellore, India

WORK EXPERIENCE

Associate Data Scientist

ANTWORKS, Bangalore

06/2020 - Present

AntWorks' is an AI-based RPA software product designed to understand structured as well as semi-[unstructured data](#) types such as handwriting and forms.

Intern - Cognitive Machine Learning

ANTWORKS, Bangalore

12/2019 - 05/2020

Signature Detection and verification

- Detection of signature is developed using object detection model of FASTER-RCNN with TENSORFLOW library
- One-shot learning with Siamese neural networks is used for verification and identification of signatures
- Applied TRANSFER LEARNING technique to generate best model with comparatively small amount of data collected from client and as well as outside sources

EDUCATION

B.TECH, Computer Science

SASTRA, Thanjavur

06/2016 - 06/2020

CGPA 8.0

Intermediate, MPC

Sri Chaitantya, Nellore

05/2013 - 06/2015

percentage 97.4

SKILLS

Python

Machine Learning

NLP

Statistical Modelling

Deep Learning

Flask

Spacy

Tensorflow

PROJECTS

Classification Module of Business Document types

- Instead of following traditional Document Classification model we developed a System where we incorporate knowledge of patterns trained by user by following Closed Domain Question and Answering Model
- Achieved maximum accuracy of 94% by classifying Mostly relevant paragraphs by learning important features in patterns provided by client instead of classifying entire document
- Provided training module for training new patterns to store in to our knowledge base required for future processes

Extracting Essential Entities from Document

- Extraction of data from images using TESSERACT and improved its accuracy by applying pre-processing techniques like removing Junk characters etc.
- Person names, addresses are extracted using SPACY module and Improved accuracy by using post processing techniques.
- Developed multiline address extraction by using traditional programming and uses SPACY , PYAP for further validation.

Sentence Comparison API

- API is developed by using FLASK which outputs similarity of input sentences.
- Provided various option to user by passing arguments to select various text processing techniques like TF_IDF , WORD2VEC, GLOVE and GOOGLE SENTENCE ENCODER.
- For Numeric datatype used REGEX to extract numeric data to compare between the inputs.

Table Extraction Module

- A generic module that covers detection of variety of tables and extract unstructured enterprise contents to a serialized form.
- Mutli page table extraction , Data Validation is also covered based on traditional programming.
- The architecture is a blend of image processing algorithms and traditional engineering techniques.

CERTIFICATES

Machine Learning course

Completed Andrew Ng Machine learning course on Coursera which teaches Basics of All Machine learning Algorithms