## M Shameem K

#### Senior Data Scientist | M.Tech

**+91-8884091353** 

- @ muhammedshameem198@gmail.com
- https://in.linkedin.com/in/m-shameem-k-23b805141
  https://github.com/shameem198
- Pengaluru, India

#### **SUMMARY**

Enthusiastic data scientist with more than four years of hands-on experience, committed to delivering cutting-edge business intelligence solutions. Proficient in advanced statistical algorithms, machine learning, natural language processing, computer vision, and LLMs. Demonstrated leadership in guiding the development of scalable and high-performance applications.

#### **EXPERIENCE**

#### Senior Data Scientist - MTS-II

#### Alphastream.ai

- 🗰 09/2019 Present P Bengaluru, India
- Managed a team of five individuals responsible for intelligent document processing with the help of **NLP** and **computer vision**.
- Launched a document understanding system utilizing computer vision technology, consistently surpassing the accuracy of GCP, Azure, AWS, and other similar platforms.
- Designed, implemented and deployed a distributed computing platform for Al-Workspace for demonstrations.
- Developed a custom document classifier for real-time processing of company-published documents, significantly automated the pipeline simulation process to 95%.
- Managed a team of three individuals to develop a data masking solution, which **enabled clients to securely process sensitive documents** such as financial and bank statements through the data pipeline with high accuracy.
- Designed, implemented and deployed AutoML solution for ML algorithms. This solution empowered financial analysts to annotate, re-train, and deploy models without any technical support.
- Designed and implemented parsers for the extraction of tables, paragraphs, and charts from HTML/PDF documents, resulting in a 50x improvement in the document onboarding process for Fitch ratings.
- Successfully designed and implemented a quick POC of end to end KYC verification with entity extraction and signature verification for Anand Rathi.
- Designed and implemented custom form parser with 99% accuracy for
- Designed and implemented a custom data point (100+) extractor from NDR and Grants documents specifically for SFC (Securities and Futures
- Implemented page classification at the beginning of the pipeline, achieved 10fold reduction in processing time for large documents, while maintaining close to 100% recall rate.
- Played a major role in both designing and implementing the FactStream data

#### Data Science Intern

#### Zwayam Pvt. Ltd.

- **=** 02/2019 08/2019 Pengaluru, India
- Effectively executed job selection bias analysis through the Apriori algorithm, leading to the detection and reduction of concealed biases in the candidate selection process.
- Successfully implemented fit/unfit resume for job description using BERT, resulting in approximately 95% accuracy and a significant decrease in process efforts.
- Enhanced resume parsing by extracting semantically meaningful sentences using SpaCy, resulting in a 10% enhancement in parsing accuracy.

#### **TECH STACK**

#### **Programming Languages**

**Python** SQL

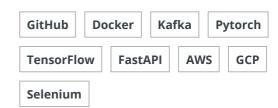
#### **Technical**

**System Design** ML/DL NLP **Computer Vision** LLM **Cloud Computing Rest Endpoints** 

#### NLP/CV

Stanford CoreNLP **NLTK SpaCy** scikit-learn HuggingFace **AllenNLP OpenAl** LangChain LlamaIndex **OpenCV MMDet** 

#### **Tools**



#### **EDUCATION**

#### M.Tech | Computational Linguistics

**Government Engineering** College, SKP

GPA **8.5** / 10.0

**=** 06/2017 - 06/2019 Palakkad, India

B.E | Computer Science and Engineering

GPA

7.0 / 10.0

#### **AMC Engineering**

**=** 06/2013 - 06/2017

P Bengaluru, India

#### **PROJECTS**

#### Supervised word sense disambiguation

- Identified the actual meaning of word within the context.
- Leveraged **SVM** and **Naive Bayes** alogirhms, Achieved around **88% accuracy** in Malayalam (native language in India).

#### Math word problem solver

- Generated mathematical representation for a descriptive mathematical word problems.
- Experimented Bi-LSTM and GRU architectures.
- Assessed which RNN architectures was more effective for the specific NLP task at hand
- **Bi-LSTM** model achieved an impressive **accuracy rate of 90%**, which outperformed previous benchmarks.

#### **INDUSTRY EXPERTISE**

**Critical Thinking** 

Leadership

## **PUBLICATIONS**

## $Automatic \ Hybrid \ Mathematical \ Word \ Problem \ Solver$

**ICSEE** 

**=** 2019

Muhammed Shameem K, Irshad M https://papers.srn.com/sol3/papers.cfm?abstract\_id=3441326

#### A Survey on Math Word Problem Solvers

#### **IJARCCE**

**並** 2019

Muhammed Shameem K, Irshad M

#### TRAINING / COURSES

#### Natural Language Processing

By Prof. Pawan Goyal | IIT Kharagpur

#### **Deep Learning**

By Prof. Prabir Kumar Biswas | IIT Kharagpur

#### **Modern Computer Vision**

By Prof. A.N. Rajagopalan | IIT Madras

# Introduction to analysis and business intelligence tools

Google Cloud Platform

#### **KEY ACCOMPLISHMENTS**



## Employee of the year - 2021 [Alphastream.ai]

- In recognition of outstanding achievements and significant contributions in document intelligence processing.
- Honored with the "Employee of the Month" award on several occasions.



Secured top15 in ICDAR 2023 Competition on Robust Layout Segmentation in Corporate Documents

https://arxiv.org/pdf/2305.14962.pdf



**Qualified GATE 2017** 

#### **PASSIONS**



Researching



**Travelling**