



Machine Learning Engineer - OCR

Join the dynamic and collaborative team at Katalyst Data Management (Katalyst)! Katalyst is seeking a Machine Learning Engineer - OCR with great problem-solving skills who is comfortable working with minimal supervision. This is an exciting opportunity to play an integral role in achieving the company's goals. Qualified candidates will bring a strong knowledge of supervised image classification and open source technologies to help us address our customer's needs.

- **Located in Canada**
- **Nominally 8.00am – 5pm Monday to Friday with occasional night and weekend work required**
- **Full-time position**

The Company

Katalyst Data Management (Katalyst) is the global leader in hosted subsurface data management solutions for the oil and gas industry. From data capture and verification, to data storage and organization, to marketing data online, our specialized data services cater to large multinational corporations, national governments and small independents. Katalyst's web-based iGlass services and their related controls, including system redundancy, are key differentiators in providing and maintaining high availability, 24/7 access for customers. Innovation and commitment has enabled Katalyst to provide excellence in both technologies and service qualities.

Key Responsibilities and Accountabilities

The Machine Learning Engineer - OCR brings a strong, proven set of AI/ML development skills to solve complex problems requiring a detailed knowledge of document scanning, optical character recognition and data extraction, technologies, tools and techniques. In this role you will help design, create and maintain our suite of internal tools supporting our service delivery to our customers. Reporting to the Chief Data Scientist and working under the guidance of the Chief Architect with other Developers, QA and DBA teams to resolve issues quickly and efficiently. Great organizational ability, a high level of attention to detail, a deep understanding of OCR and a natural curiosity are also required for this role.

Key Responsibilities

- Understand the requirements and objectives of the assigned projects
- Recommend appropriate approaches and technologies to deploy to achieve the objectives
- Assist in the deployment of approved technologies and tools
- Collaborate with colleagues to select and produce appropriate data sets
- Design the development platform and approach to the solution
- Iterate the development process to improve effectiveness of the models
- Work with the Chief Architect to productize your solutions
- Train Operations staff in the use of your solutions.
- Monitor production processes and continue to iterate improvements

Knowledge & Skills Required

- Computer vision: text recognition and machine vision
- Natural Language Processing, content extraction, classification, machine translation, text generation
- OpenCV, Tesseract, TensorFlow and similar open source technologies
- Python, Pytorch, SQL, Sagemaker, Numpy, Pandas, Keras, Random Forest, XGBoost and similar
- Experience with supervised image classification techniques
- Familiar with the Linux OS (CentOS preferred)

Preferred Skillsets

- Ability to work in a self-directed manner
- Curious mind, driven to solve complex problems spanning multiple technologies through coding of utilities and applications
- Deep understanding of document scanning, OCR and process automation
- Researcher's mindset, able to maintain motivation after setbacks and continue to make progress
- Experience with object detection, long short-term memory (LSTM) networks and recurrent neural networks (RNN)

Required Education and Experience

- 3+ years of relevant software development experience
- Bachelor's in computer science or related quantitative field
- Demonstrated experience and success in computer vision and text extraction projects

Occasional travel may be necessary; however, most of the work can be performed from the individual's home or office remote from the Katalyst physical office.

Applications can be emailed in confidence to hr-cal@katalystdm.com.