

Case Study

Streamlined MLOps for U.S.-based Real Estate Consulting firm



Client Overview

- U.S.-based Real Estate, Infrastructure, and Environmental Consulting services firm that helps real estate property owners – both public and private – design efficiently, make better informed decisions about infrastructure projects, and cut project implementation costs.

Business Challenges / Need

- Automate information of interest extraction from survey text present in 1000s of land deed documents
- The information required was in the form of a dozen entities like distance, direction, and curve details for each traverse
- It was imperative that the text extraction ML model be trained with good quality annotated data for higher accuracy
- Its key to scale data annotation process, while keeping costs in check

Solutions

- The land deed documents were staged in AWS S3, which were retrieved and passed through the entity recognition model
- Traverse entities such as distance, direction, curve were persisted in AWS RDS, which served as a lookup for the data annotation process
- LSTM CRF model was leveraged to identify and annotate 12 entities from 2000+ deed documents
- The annotated documents were then converted into XML format, which can be fed directly into the ML model
- Data annotation process was streamlined using GATE software for cost efficiency
- The MLOps solution was containerized with AWS ECR and deployed on AWS EC2

Technology Stack

- AWS S3, AWS RDS, AWS ECR, AWS EC2, GATE software, Named entity recognition, LSTM-CRF

Business Impact

- ML model trained using the tagged data achieved **85% F1 score**
- **30% lower time to annotate** from streamlining the entire process with GATE software
- **Lower data annotation cost** was realized through streamlined process and effective quality control mechanism that minimized human errors