CSCI 5410: Assignment 4

Part A: Summary on AWS Sagemaker and Comprehend

SageMaker [1]

The AWS SageMaker is basically a fully managed end-to-end Machine Learning (ML) service provided by Amazon to build, train and deploy ML models. The biggest take-away from AWS SageMaker is that it is very efficient and seamless with immense support of SDKs and other AWS provided services. The SageMaker allows the users to build highly accurate models that allows it to improve overtime with the fed data and the decisions that it makes. The developers does not have to do any heavy lifting with managing ML environments and infrastructure. Below I have mentioned in detail how AWS SageMaker makes the ML models;

1. The very first thing that needs to be done is feed lots of data (metadata) to the SageMaker ML model.
    a. There is seamless connection between the AWS S3 and AWS Redshift.
2. After the data is uploaded, then that data is used to train the ML model.
3. Now, to maximize the signal and reduce the noise in data it is needed to convert the raw metadata and convert it into Features through a process known as Feature Engineering.
    a. SageMaker Data Wrangler can be used to convert, combine and transform raw tabular data into meaningful features needed to train the ML model efficiently.
4. The SageMaker lets us check-in and check-out various features along with creating multiple version of them.
5. The SageMaker also has features where we can add description to the features and also assign tags that assist in searching features making it easy to use them in future.
6. Great models can be made if they are trained on balanced set of Data and Features, which can be achieved by using SageMaker Clarify.
7. The various tools provided by SageMaker are;
    a. Visual Editors
    b. Debuggers
    c. Profilers
    d. CI/CD for ML
8. AWS SageMaker also has the SageMaker Notebooks which are basically Jupyter Notebooks.
    a. So it is like a manager server to help us in using the inside of AWS.
9. From UI perspective there are three things that AWS SageMaker has;
    a. AWS Console
    b. SageMaker Notebooks and,
    c. SageMaker Studio

According to me the AWS SageMaker can be used in the following ways in a Car-Rental Application:
In a Car-Rental Application the main goal or insight is to understand which type of car, which time of the year, the region and even the age group that rent cars the most. These are the kind of insights that will drive the business and assist in making informed decisions. I will utilize AWS SageMaker's service to convert my Car-Rental Application's raw metadata to usefull Features like what age group prefers which type of car or which type of car is preferred for which holiday locations etc. I will also use the SageMaker Data Wrangler to train my model along with using SageMaker Clarify to maintain a balanced set of data and features extracted form the Car-Rental Application.

Comprehend [2]

The AWS Comprehend is simply a very powerful service that uses Natural Language Processing (NLP) to extract insights from the contents of the documents. There is a wealth of untapped information present in the documents which could be analysed to gain valuable insights. In reality the AWS Corehend uses various ML models that help in reading the contents and analyse them. The best part is that the user does not have to worry about building, training or even maintaining the ML model as the AWS Comprehend takes care of all those things. The AWS Comprehend is capable of processing any text files that are in the UTF-8 format or even semi-structured documents like the PDF or Word documents. The applications of AWS Comprehend are enormous in real-life scenarios as we have been generating semi and un-structured data which hold key insights helping in taking informed business decisions. Below are some of the insights that AWS Comprehend develops after processing any document; Entities, Key Phrases, PII, Language, Sentiment and the most important Syntax. There is a unique feature provided by AWS Comprehend known as the Comprehend Custom that helps the user to customize their specific requirements and build ML-based NLP models by using AutoML. Another interesting thing about AWS Comprehend is that it used pre-trained model to examine and analyse the documents and it also continuously trains itself on a large set of data without any requirement of training data. In AWS Comprehend there are various ways to process documents like Single-Document Processing, Multiple Document Synchronous Processing and Asynchronous Batch Processing.

According to me the AWS Comprehend can be used in the following ways in a Car-Rental Application:
In any application there is a lot of data that is being generated and stored by the application like for the Car-Rental Application the data is User information, Car information, Car registration information, Rental information and logs, Car maintenance reports, Car vendors information, Chatbot history or query resolution information or log, social media posts, emails and much more. A lot of this information is stored in the form of scanned copies of document and a NoSQL database (basically unstructured data). The Comprehend will basically help me read those data and give me extracted data and topics with a confidence score helping me in taking informed decisions.

References

[1] Amazon, "Amazon SageMaker Documentation," Amazon Web Services, 2021. [Online]. Available: https://docs.aws.amazon.com/sagemaker/?id=docs_gateway.

[2] Amazon, "Amazon Comprehend Documentation," Amazon Web Services, 2021. [Online]. Available: https://docs.aws.amazon.com/comprehend/?id=docs_gateway.