Evaluation Form - Technical Background Review

Student Name: _	Shreyas Mhasawade
Project Advisor:	Xiaoli Ma
Team Name:	ViaSat AMEND- Machine Learning
Team Members:	Tyler Cole, Adrija Bhattacharya, Mikias Balkew, Chris Rothmann

/ 100	Total - Technical Review Paper
	Style, readability, audience appropriateness, conformance to standards
	 Grammar, spelling, punctuation
	Organization
	Clear flow of information
	Introductory paragraph
/ 40	Effectiveness of Writing, Organization, and Development of Content
	Reference list in proper format
	 Appropriate citations in body of text
	Quality of the sources
	 Sufficient number of source types (at least four)
	 Appropriate number of sources (at least six)
/ 30	Use of Technical Reference Sources
	Overall quality of the technical summary
	 Implementation of the technology
	 Underlying technology
	 Current state-of-the-art and commercial products
/ 30	Technical Content

Machine Learning for Real-Time Systems

Introduction

As the number of satellites in low Earth Orbit has increased, it has become more important for communication dishes on Earth to keep continuous contact with their respective satellites. It can be a painstaking process for the dish to find and lock onto the satellite that was lost manually. Real-time machine learning utilizes special techniques to parse raw input data and teach itself without requiring human assistance [1]. This paper will investigate real-time machine learning and will examine both how a real-time machine learning model can be developed and how it has been applied.

Applications

Tracking Heart Disease

Researchers have proposed using real time machine learning to detect early-stage heart disease in patients [2]. In this proposal, researchers created a model which utilizes Apache Spark Streaming to synthesize data from multiple socket sources, including wearable medical devices and healthcare databases. The data is processed through the machine learning model and relevant data is fed back into the model to improve predictions. The database input data contains records of both individuals with heart disease and individuals without it. By using different health attributes from members of both categories, the model performs calculations and can generate key insights as output.

The work was carried out on a Linux machine equipped with a core i7 processor and 8 gigabytes of RAM. The relatively low-cost equipment used to carry out this project demonstrates that the majority of cost in implementing real time machine learning comes in the time spent creating the model itself. The model does not require advanced hardware to run. By using the Spark and Cassandra models to build upon their technology, these researchers emphasized creating a scalable model that can be used for increasing amounts of data as the platform grows.

Technology behind Machine Learning

Traditional vs. Real-Time

As opposed to traditional, *supervised*, machine learning, real-time machine learning is *unsupervised* [3]. This means that it uses input data that is unlabeled. In a traditional setup, a human with domain expertise will have to manually label datasets that are fed into the model. The model then takes the input data and processes it accordingly to identify relationships between certain variables. Supervised

learning models are usually more accurate, but they can only be trained offline, after a human has put in time to make the data readable for the machine [4].

On the other hand, unsupervised learning means that the during runtime, the machine takes in raw, unlabeled data, and processes it the best way it can. Although initial performance for a real-time system will not be as accurate as a traditional one, the real-time system will be able to run without requiring human assistance, and the model will never have to be shut down for additional model training [5].

Real-time Techniques

Real-time systems process unfamiliar data by using different techniques, including clustering, association, and dimensional reduction [5].

Clustering is when raw input data points are grouped together and represented by patterns and structures rather than by labels. Different types of clustering will group data points in different ways, with some mandating that a data point must belong to a single cluster and others allowing data points to be in contact with multiple groups. Eventually, after many cycles of data processing, the clustering techniques all normalize, and the data is grouped accurately [5].

Another technique real-time systems utilize is association. The model will use multiple "rules" to see if a particular data point has a relationship with another. By observing multiple different parts of a particular data point, the model will use "rules" to see if any other data points have matching data attributes [5].

One more technique a real-time system will utilize is dimensional reduction. This is when the model will reduce a data set's complexity or size in order to process it more quickly. Since all calculation is happening in real-time, a system might not have the processing capability to handle all the data that is being fed. Traditional models are offline, so they have the luxury of spending additional processing time while reading new data, but a real-time model needs to find a way to improve its performance. Reducing data complexity can allow a model to perform better and output readable data [5].

Implementation Requirements

Although there are many different approaches to begin building a real-time model, many approaches follow similar processes. The first thing that needs to be created is a place for the model to be stored. Some models do this by utilizing NoSQL Databases in the cloud, and some do this by storing them locally until release, at which time they deploy the model on their systems [6]. Real-time models also require a framework that can deal with large amounts of input data. Some of these frameworks include Apache Spark and TensorFlow [7]. Developers can utilize these frameworks and their built-in data manipulation functions in order to build models that fit their use-cases.

References

- [1] Hazelcast, "What is Real-time Machine Learning?" Mar. 10, 2020. [Online]. Available: https://hazelcast.com/glossary/real-time-machine-learning/. [Accessed Oct. 6, 2021].
- [2] A. Ed-Daoudy and K. Maalmi, "Real-time machine learning for early detection of heart disease using big data approach," 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), 2019, pp. 1-5, doi: 10.1109/WITS.2019.8723839.
- [3] J. Delua, "Supervised vs. Unsupervised Learning: What's the Difference?" Mar. 12, 2021.
 [Online]. Available: https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning.
 [Accessed Oct. 7, 2021].
- [4] IBM Cloud Education, "Supervised Learning," Aug. 19, 2020. [Online]. Available: https://www.ibm.com/cloud/learn/supervised-learning#toc-unsupervis-Fo3jDcmY. [Accessed Oct. 7, 2021].
- IBM Cloud Education, "Unsupervised Learning," Sep. 21, 2020. [Online]. Available: https://www.ibm.com/cloud/learn/unsupervised-learning. [Accessed Oct. 7, 2021].
- [6] B. Weber, "Approaches for Building Real-Time ML Systems." Sep. 6, 2021. [Online]. Available: https://towardsdatascience.com/approaches-for-building-real-time-ml-systems-79ea0e340269.
 [Accessed Oct. 7, 2021].
- [7] M. Todi, "Real-Time Machine Learning," Oct. 18, 2019. [Online]. Available: https://medium.com/analytics-vidhya/real-time-machine-learning-7aa55dafb2b. [Accessed Oct. 7, 2021].