Abstract

In today's data-driven world, data streaming has emerged as an essential element for businesses worldwide.

# Edris Safari

## DSC680 Project2, Milestone1

Project Proposal

Real-Time Big Data Streaming with Kafka

# Topic

# Business Problem

# Datasets

# Methods

# Ethical Considerations

# Challenges/Issues

Challenges and issues that we could face in this project must be considered and addressed. This is part of risk assessment that all projects must go through. The table below lists the risks and their mitigation.

|  |  |
| --- | --- |
| Risk | Mitigation |
| Data Quality | Ensure data quality by performing a preliminary analysis |
| Data Security | Ensure data is secure both incoming and outgoing. Enable/utilize security measures. |
| Data Availability | Ensure data is available without interruption or delay that may affect the performance of the analysis |
| Technical Challenges | Enforce fault tolerance, redundancy, connection integrity |
| Ethical Violations | Ensure procedures are put in place that will reduce and remove risk of ethical violations by all parties involved. |

# References

1. <https://blog.logrocket.com/apache-kafka-real-time-data-streaming-app/>
2. <https://kafka.apache.org/intro>
3. <https://analyticshut.com/kafka-producer-and-consumer-in-python/>
4. <https://www.youtube.com/watch?v=tFlYoEJsT2k>
5. <https://www.youtube.com/watch?v=LjjPjT6R9Bg>
6. https://medium.com/@alipazaga07/setting-up-your-first-apache-kafka-development-environment-in-google-cloud-in-15-minutes-ffdb1623e125
7. <https://www.learningjournal.guru/courses/kafka/kafka-foundation-training/kafka-in-gcp/>
8. <https://www.youtube.com/watch?v=RYC-7wECMds&list=PLa7VYi0yPIH14oEOfwbcE9_gM5lOZ4ICN>
9. <https://spark.apache.org/docs/latest/structured-streaming-programming-guide.html>
10. https://phoenixnap.com/kb/install-spark-on-ubuntu
11. François Chollet. Deep Learning with Python (Kindle Locations 1504-1508). Manning Publications Co.. Kindle Edition.