Course Information

Course: Cloud Computing DATA/MSML 650

Term: Fall 2025

Credits: 3

Course Time:

• PCS1: Monday 4:00–6:45 PM

Instructor Information

Instructor: Dr. Samet Ayhan

Project Proposal: Real-Time Economic Forecasting Using Alternative Data

1. Group Members

Akul Ukey
Aman Kumar Sahu
Ankur Sheth
Sahil Parab
Supriya Nagnath Kadam

2. Planned Contributions

To effectively manage the project's complexity, we have outlined the core responsibilities for each member:

Akul Ukey: Will design the overall AWS architecture, set up IAM roles and permissions, and ensure secure networking. Akul will also orchestrate the final end-to-end pipeline using AWS Step Functions and SageMaker Pipelines, managing the integration between all components.

Aman Kumar Sahu: Will create the data ingestion pipelines for our various data sources (News APIs, AIS Shipping Data, FRED Economic Data). Aman will also set up and manage the S3 data lake structure and use AWS Glue to catalog and prepare the raw data for analysis.

Ankur Sheth: Will lead the development of the computer vision feature engineering pipeline. This includes setting up the labeling job in Amazon SageMaker Ground Truth, as well as training, tuning, and deploying the object detection model on SageMaker to process the satellite images at scale.

Sahil Parab: Will develop the news sentiment analysis pipeline using Amazon Comprehend and process the AIS shipping data with Amazon Athena. Sahil will also manage the critical data fusion phase, using Amazon EMR (Spark) to join the engineered features from all pipelines into a final, unified dataset.

Supriya Nagnath Kadam: Will focus on training and evaluating the final forecasting model using services like Amazon SageMaker Forecast or custom algorithms. Supriya will also deploy the finalized model, set up the automated inference pipeline, and build the project dashboard in Amazon QuickSight.

3. Problem Statement

Traditional macroeconomic indicators, such as monthly retail sales figures or manufacturing output, are lagging indicators. They are often released weeks or even months after the period they describe. This time lag poses a significant challenge for investors, policymakers, and business leaders who need to make timely decisions based on the current state of the economy. Our project aims to solve this problem by developing a system that can predict these key indicators in near real-time, providing a valuable 'early warning' signal of economic trends.

4. High-Level Approach and Novelty

Our approach is to create a sophisticated forecasting model that synthesizes data from multiple unconventional, high-frequency 'alternative data' sources. Instead of relying on a single data stream, we will fuse signals from three distinct domains:

1. Commercial Activity: Using computer vision on satellite imagery to measure activity at major retail centers.
2. Market Sentiment: Using Natural Language Processing (NLP) on financial news to gauge economic optimism or pessimism.
3. Global Trade Flow: Using maritime GPS data to quantify the movement of goods into major ports.

The novelty of our project lies in the fusion of these disparate, large-scale datasets within a fully automated, cloud-native pipeline. We are moving beyond traditional econometrics to build a multi-modal AI system that learns from a richer, more timely, and more holistic view of economic activity.

5. Implementation Tools

We will leverage a comprehensive suite of AWS services to build our platform:

Data Storage & Lakehouse: Amazon S3, AWS Glue (Data Catalog)
Data Ingestion & Streaming: AWS Lambda, Amazon EventBridge
Data Processing & Analytics: Amazon EMR (with Apache Spark), Amazon Athena, AWS Batch
Machine Learning & AI:
- Amazon SageMaker (for custom model training, Ground Truth, and Pipelines)
- Amazon Comprehend (for managed NLP)
Orchestration: AWS Step Functions
Visualization & Reporting: Amazon QuickSight
Core Infrastructure: IAM, Amazon VPC