

Mission: Predictable
Team Ada Lovelace

The “Map My Risk” Application: Problem Statement
Hieu Dinh, Arias Hathaway, Fiona Tan, Gabriela Trindade, Sofia Sikandar

Communities across the world have been faced with the disruption of our most basic routines with the arrival of the coronavirus pandemic. Mundane errands, like a trip to the grocery store, now carry the risk of placing individuals at the virus’s frontline and introduce a level of danger to every activity outside the home. Many applications already exist to make errands as convenient as possible, such as mapping routes to avoid traffic and checking whether a business is particularly busy at a given time. In the time of this global pandemic, applications have the opportunity to develop functionality beyond convenience and convey information that helps communities make decisions to limit the spread of the virus. Technology can guide these decisions by providing a clear picture of how COVID-19 is spreading through a given community and how individuals can socially distance in a responsible manner.

Team Ada Lovelace addresses this need with an application called Map *My Risk* that employs machine learning to inform users of their states’ current transmission rates (R_t). The transmission rate is a valuable tool to measure how the virus is spreading in a given region. The purpose of the application is to check in with local conditions before leaving the house for daily activities. The current version of Map *My Risk* monitors the pandemic’s status for residents in the United States. Upon launch, users key in their state, and the application will return the number of current cases, current transmission rate, forecasted transmission rates, forecasted new cases, and a map summary of new cases and transmission rates across the United States.

The application’s backend functions in the following way. The moment a user keys in their state, the program triggers an Amazon Lambda function to hit the Amazon SageMaker endpoint in which the machine learning model predicts the R_t value for the next 5 days in the user’s chosen state. Amazon Kinesis Data Firehose is used in our pipeline to stream consistent predictions to Map *My Risk*’s users. The raw data is prepared and transformed before being returned to users. The columns of the dates, states, transmission rate values, and number of new cases are consistently updated from the dataset on rt.live. After the data is transformed, a lambda function is triggered to load the data into S3 and then into our model in Amazon SageMaker. The dataset is consistently ingested, transformed, and delivered to give Map *My Risk*’s users the most up-to-date predictions.

The model utilizes seven, nation-wide datasets with state-level data describing how often people leave their home and for which purposes. The model learns from the supervised regression algorithm and utilizes the Light Gradient Boosting Machine framework (LGBM). The LGBM framework uses a sequential, tree-based learning algorithm. LGBM was chosen because of its capability of handling large datasets, requiring minimal GPU power and yielding high accuracy. The independent variables were dates, historical cases, location, weather, and mobility information. The dependent/forecasted variable was the transmission rate, which is the value the user receives in the Map *My Risk* application.