INTRODUCTION

J.P.Morgan…

- Operates ~ 5,000 US branches.
- Opened ~ 400 new branches between 2018 and 2022.
- Seeks optimal locations for future branches.

Studying urbanization is vital to accurately gauge an area's potential for success. Hence, this capstone aims to predict the urbanization rate across the US with fine granularity.

THE DATASET

Crops
- Forests
- Water
- Buildings

Dynamic World: Google's 10m-resolution landcover segmentation, from 2016 to present, with daily images and nine labels.

THE MODEL

We have three baselines for the prediction task: Historical Average, XGBoost, and Convolutional Neural Network.

An initial classifier partitions low and high urbanized areas. Two different regressors are used. For low-urbanized areas, we use XGBoost on aggregated features (noisy images), whereas for high urbanized areas, we use a Convolutional Long-Short Term Memory (ConvLSTM) framework for video predictions.

ConvLSTM: Capture spatio-temporal relationships using previous frames to predict future ones.

DATA EXTRATION PIPELINE

(Automated extraction from Google Earth)

FISHNET CREATION PIPELINE

(Generate fishnet with bounding box coordinates)

DATA PROCESSING PIPELINE

(Impute missing pixels, compute metrics)

DATA LOADER

(Real-time image splitting to accelerate training)

New features: Urbanization and Rates
- (Integration of urbanization and urbanization rates as new features in J.P.Morgan's downstream models.)

SUMMER 2023