

COVID-19 DAILY DEATH TOLL PREDICTION IN USA

Fall 2020 CSCI-SHU 360 Machine Learning

Veronica Hu - Yuan Huang - Yumeng Zhang





01. INTRODUCTION

Problem situation
and significance

02. DESCRIPTIVE ANALYSIS

Datasets and
interesting insights

03. MACHINE LEARNING

How we select and train
ML models



04. RESULTS, CONCLUSION

Final results and
future works

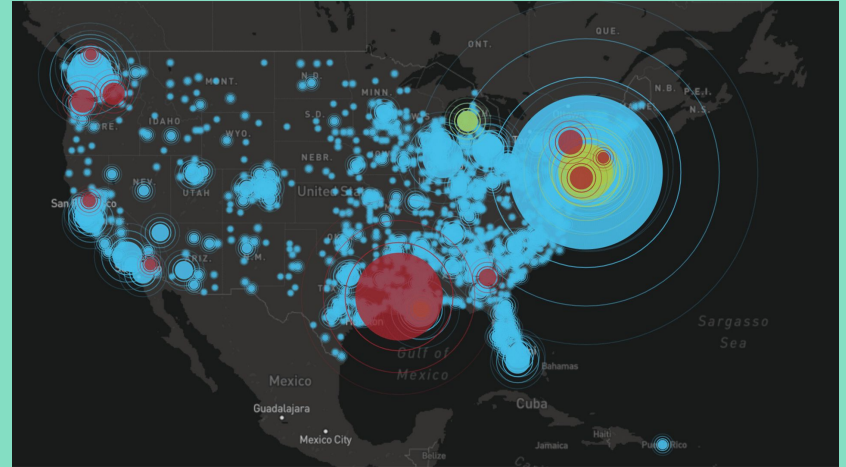
01. INTRODUCTION

1,594,204

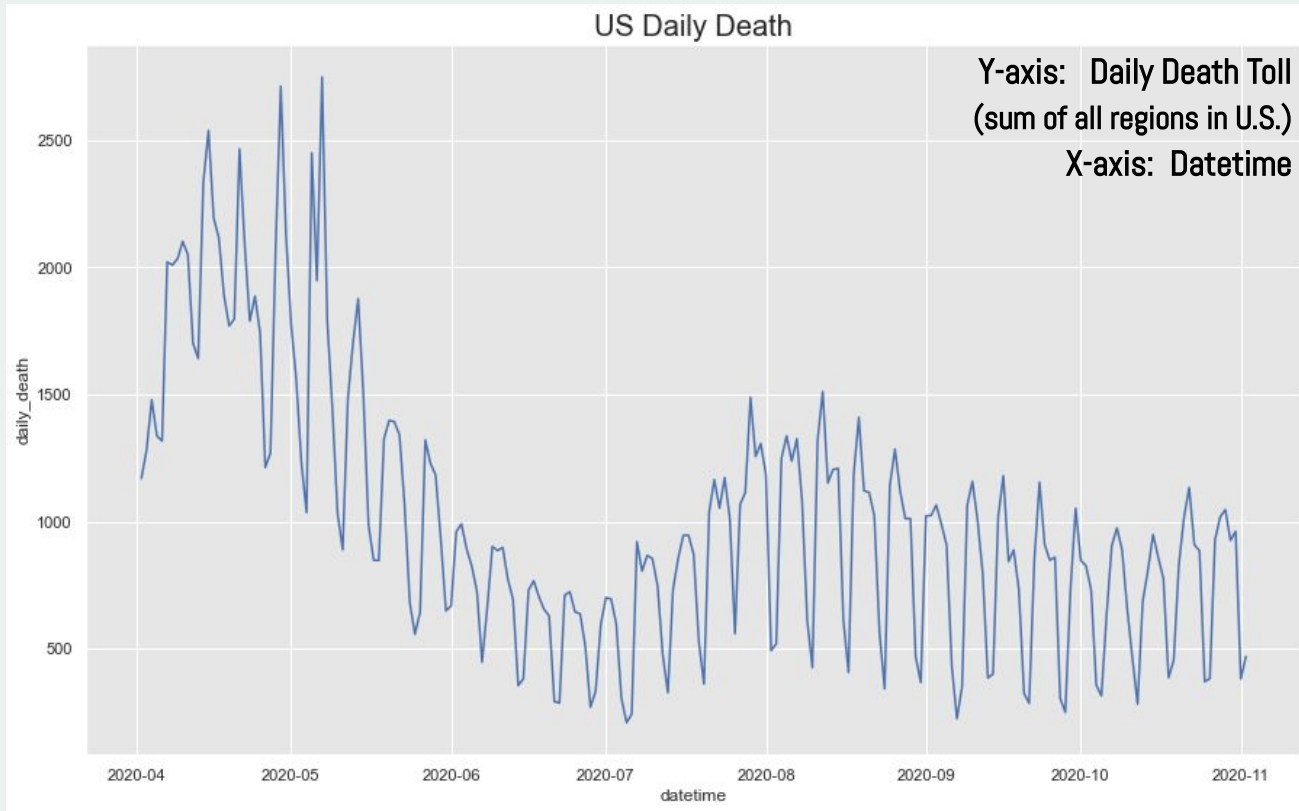
Deaths caused by COVID-19 worldwide

285,643

Deaths caused by COVID-19 in the United States



WHAT IS TIME SERIES DATA?



02. OUR DATASET

DATASET 1: GENERAL INFO

E.g. deaths, tests,
positive cases

DATASET 2: HOSPITAL CAPACITY

DATASET 3: MOBILITY

Number of population
inflows and outflows

MOBILITY DATASET VISUALIZATION



03. METHODS

A. FEATURE SHIFTING

Autoregressive Model (AR)



Predicted daily
death cases
on 10/1

= $C + b_1 \cdot$

Daily death
cases on 9/30

+ $b_2 \cdot$

Daily death
cases on 9/29

+

+ $b_k \cdot$

Daily death cases
on k days before

Shift other time-series features from previous k days

Elbow point: $k = 10$

03. METHODS

B. MODELING

01

Baseline Model:
Rolling Mean

02

SGDRegressor (Ridge)

03

Support Vector Regressor

04

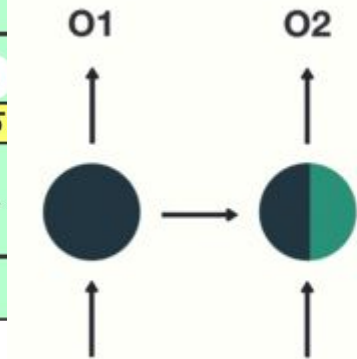
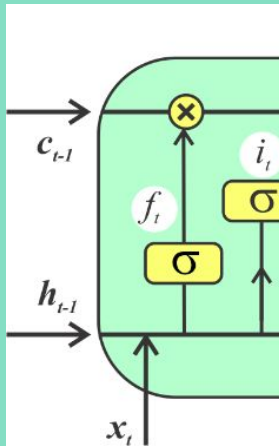
Decision Tree Regressor

- Single tree
- Random Forest
- XGBoost

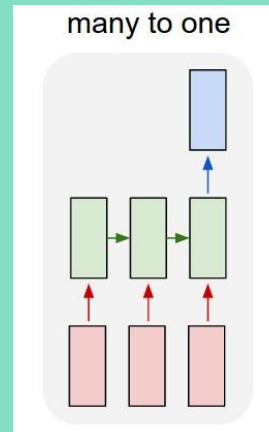
05

RNN-LSTM
(Long short-term memory)

LSTM



What time is it ?



04. FINAL RESULTS

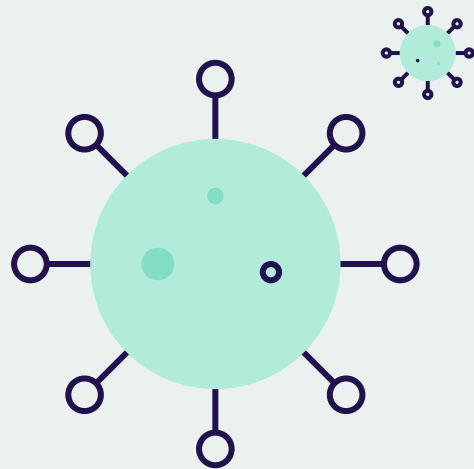
Baseline: 0.22

Colsample_bytree = 0.5,
Gamma = 0.58,
Learning_rate = 0.59,
Max_depth = 4
N_estimators = 4
Subsample = 0.95

	SGD	SVR	Decision Tree	Random Forest	XGBoost	LSTM
test_MSE	0.21	3.50	0.15	0.14	0.08	0.23
train_MSE	0.89	0.12	0.34	0.29	0.13	0.07

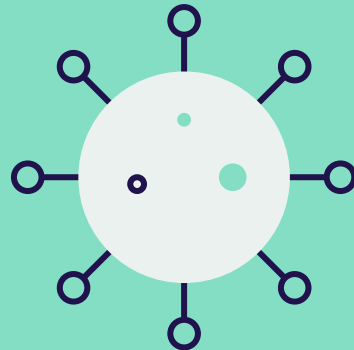
04. FUTURE WORKS

- Models
 - Better hyperparameter
 - Ensemble modeling -- stacking
 - Make Improvements for LSTM
- Dataset
 - December and future daily death data





THANK YOU!



Welcome to reach us out if you have any questions regarding our project

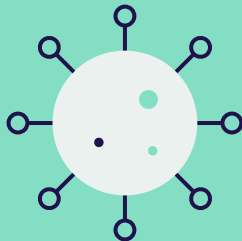
Our Github link:

https://github.com/Yuan-032/Machine_learning_final_project

Yumeng Zhang (yz5160@nyu.edu)

Veronica Hu (jh6181@nyu.edu)

Yuan Huang (yh2741@nyu.edu)



CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik