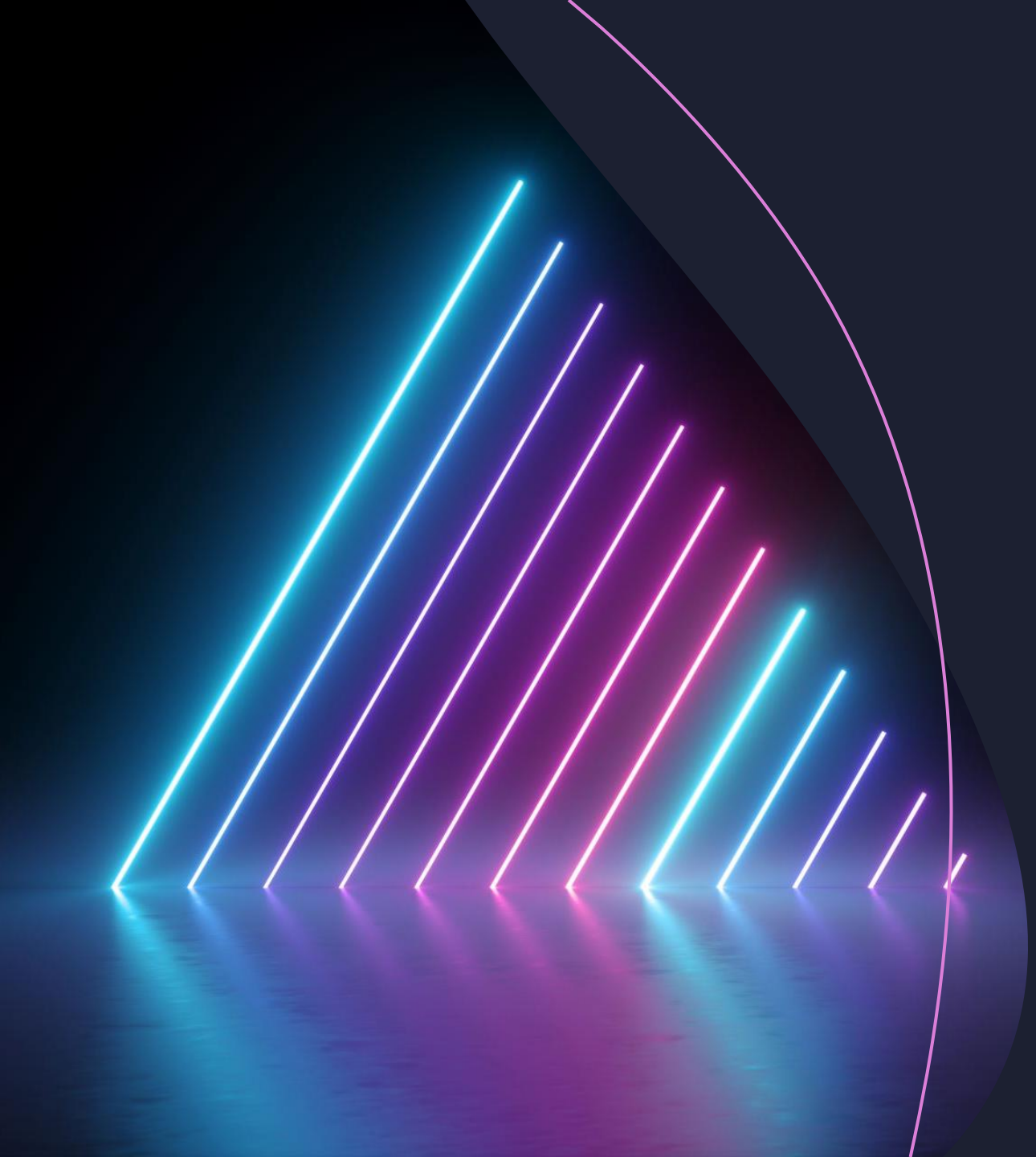


UBS Challenge

Datathon 2024

support nectar machines

Outline

- Introduction
- Data Engineering
- Model
- Results
- Outlook

Introduction

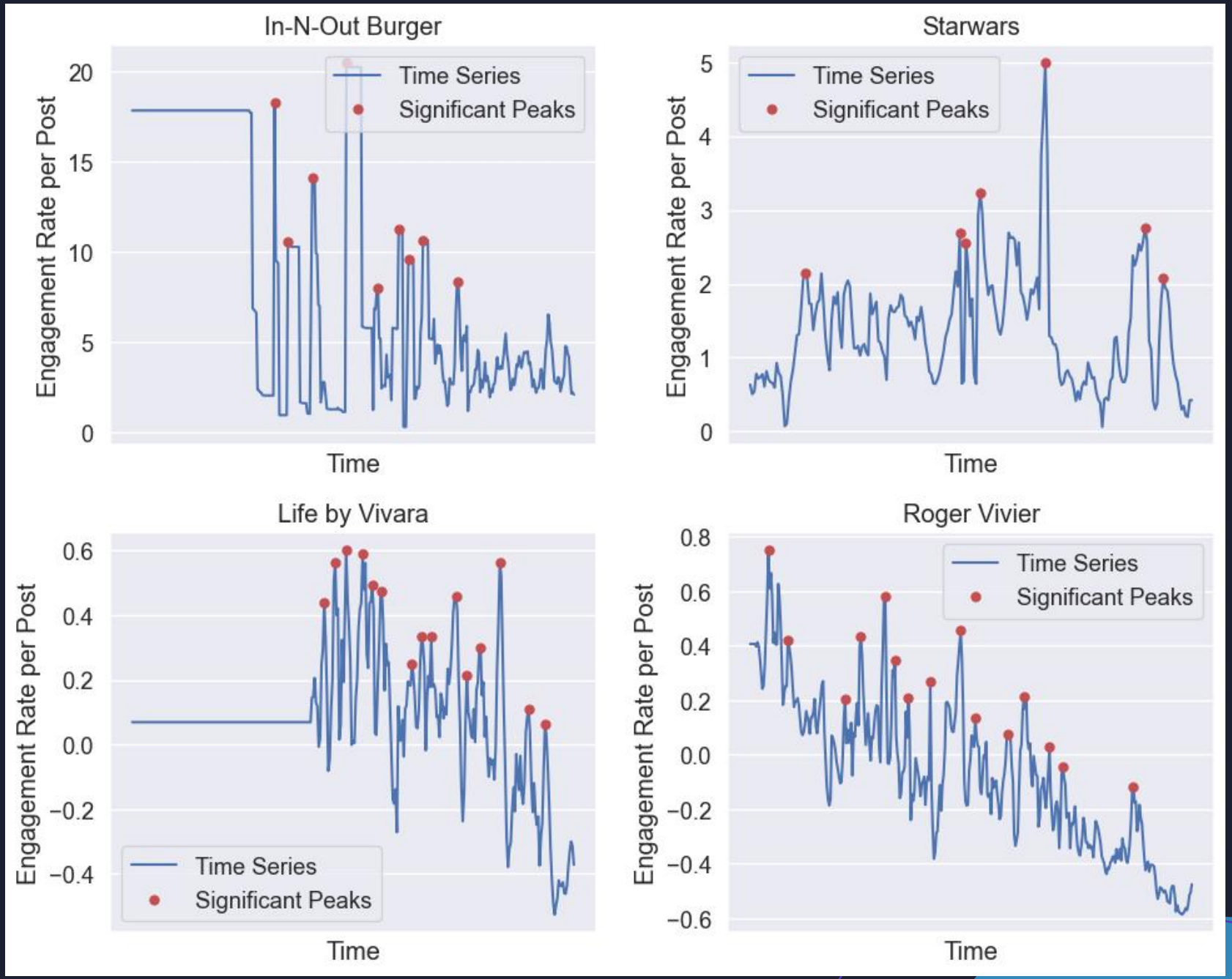
- Tracking of popularity of brands on Instagram
- Aim: Model social media interaction
- Approach:
  - Data Exploration and Engineering
  - Easy Baseline
  - Model
  - Analysis

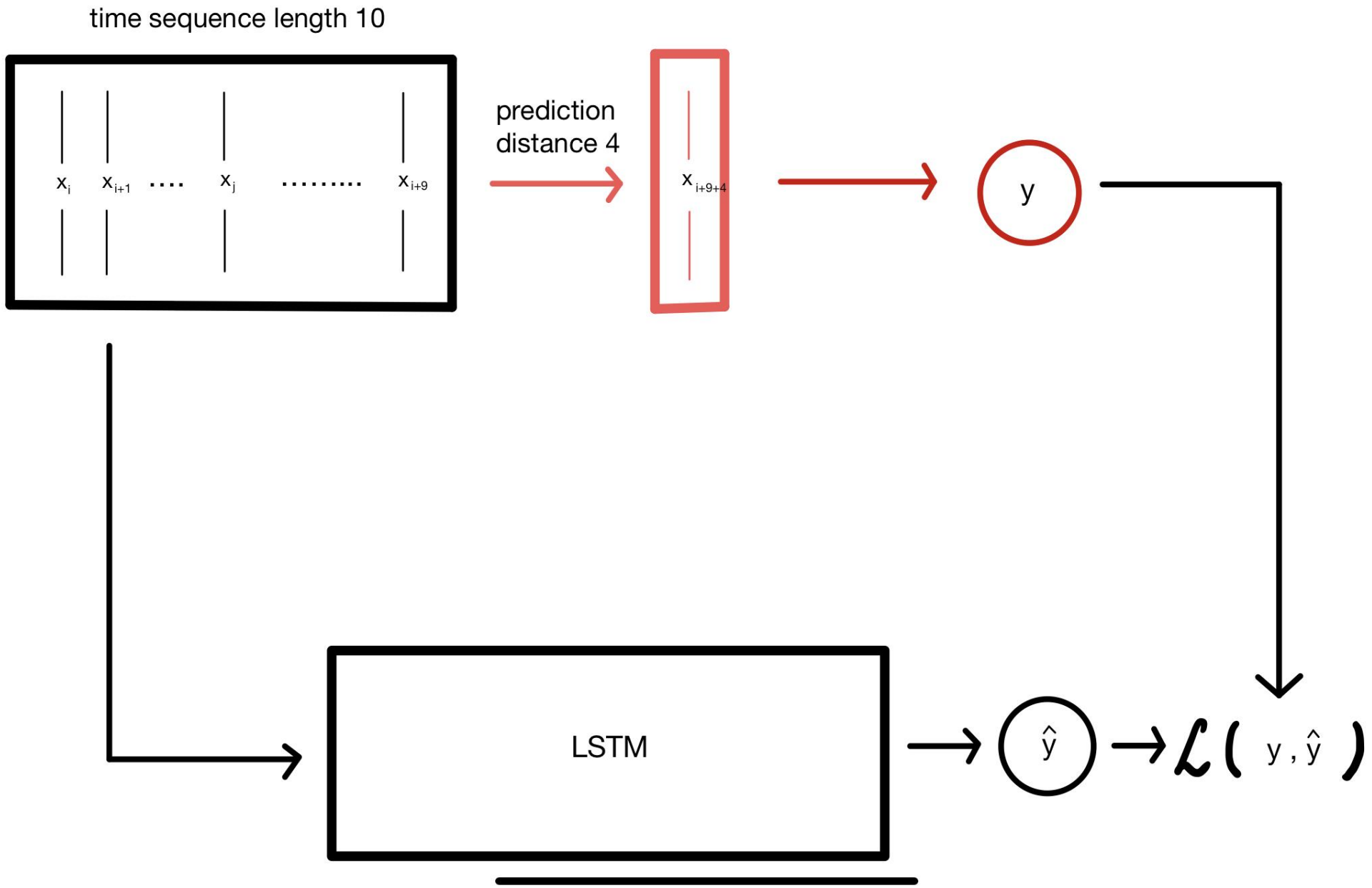
Data Engineering

- Removal of uninformative columns (e.g. calculation type)
- Standardization
- Train (80%) - Test (20%)
- Engagement rate per post: $$erpp = \frac{likes + comments}{followers * (pictures + videos)}$$

Model Approaches

1) Engagement Spike Detection Model
2) LSTM
3) Gaussian Process

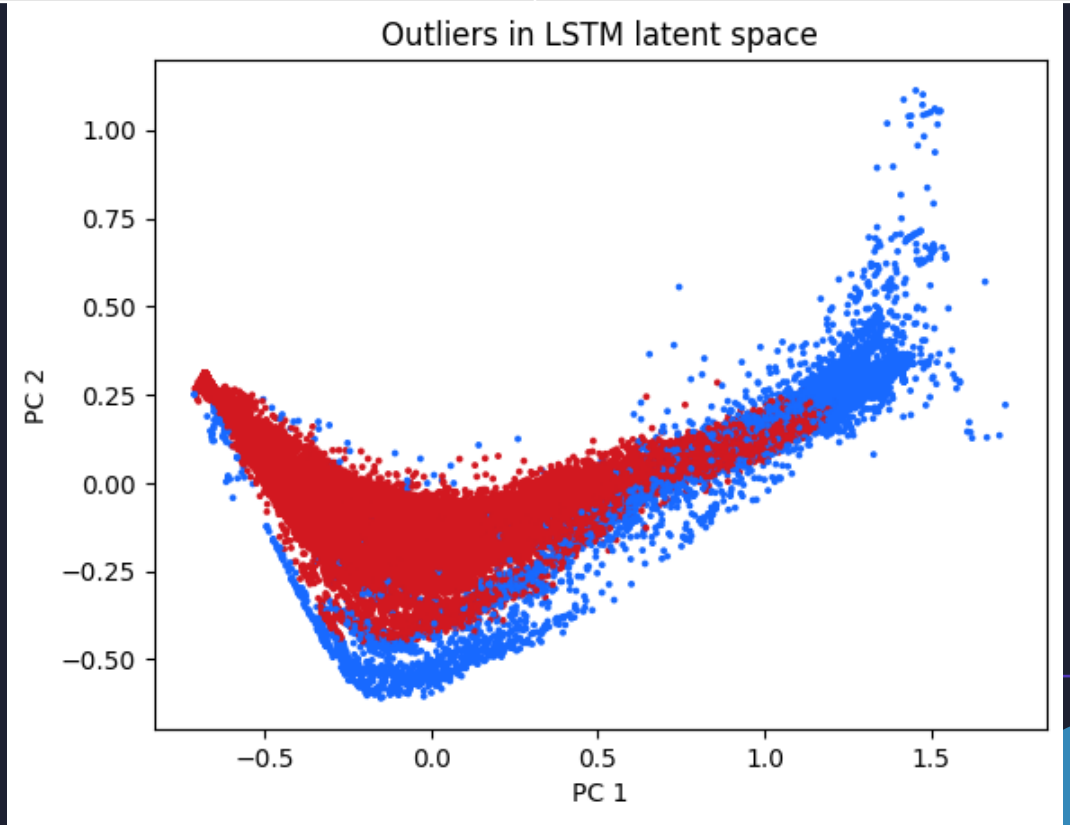

In-N-Out Burger
Engagement Rate per Post
Time
Time Series
Significant Peaks
0
5
10
15
20
Starwars
Engagement Rate per Post
Time
Time Series
Significant Peaks
0
1
2
3
4
5
Life by Vivara
Engagement Rate per Post
Time
Time Series
Significant Peaks
−0.4
−0.2
0.0
0.2
0.4
0.6
Roger Vivier
Engagement Rate per Post
Time
Time Series
Significant Peaks
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8


time sequence length 10
x_i x_{i+1} .... x_j ..... x_{i+9}
prediction
distance 4
x_{i+9+4}
y
LSTM
$\hat{y}$
$\mathcal{L}(y, \hat{y})$

Results 1

Mean Squared Error:

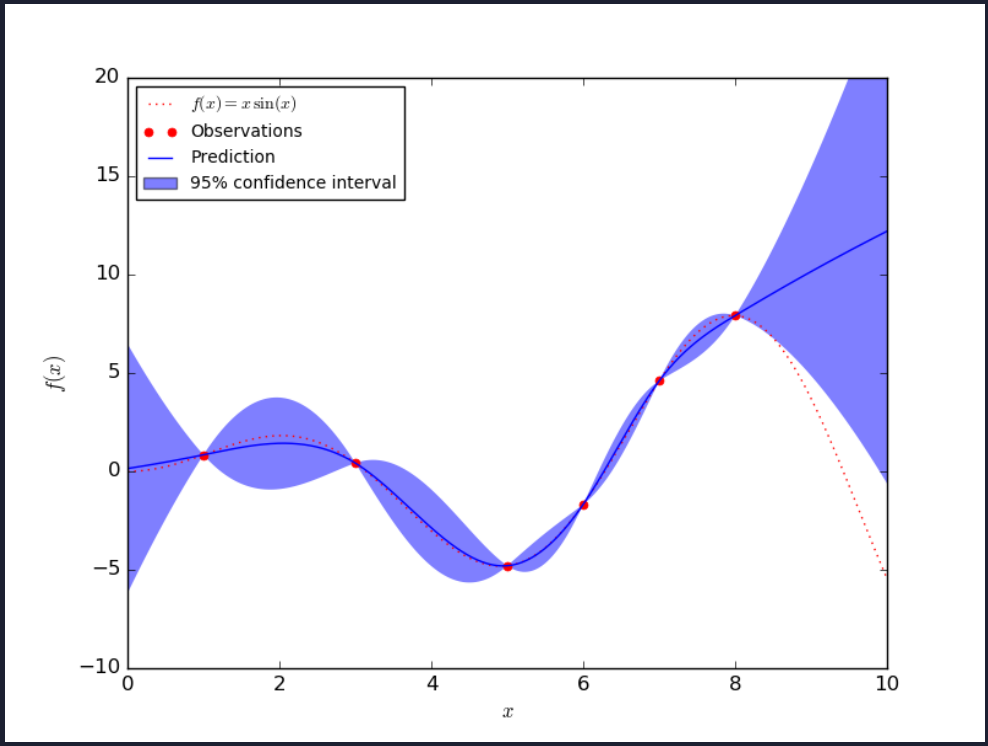
Easy Baseline (Linear Regression)	Our Model (LSTM)
1.21	0.78



Results 2

Mean Squared Error:

Our Model (LSTM)	Gaussian Process
0.78	0.26



- Allows us to estimate uncertainty of future prediction, essential for investment decisions
- Highly accurate method for current data
- As data becomes more sophisticated it remains to be seen which model is more viable

Outlook

- Improve training objective
- Investigate degrees of freedom in latent space of LSTM
- Use model insights for recommender system

Thank you very much

David, Frederieke, Arvid, Yi-Yi

By support nectar machines