SIGIR 2020: Finir

Rank1

Team Name : maimang

Respondent : Weilong Chen

SIGIR 2020
XI'AN·CHINA

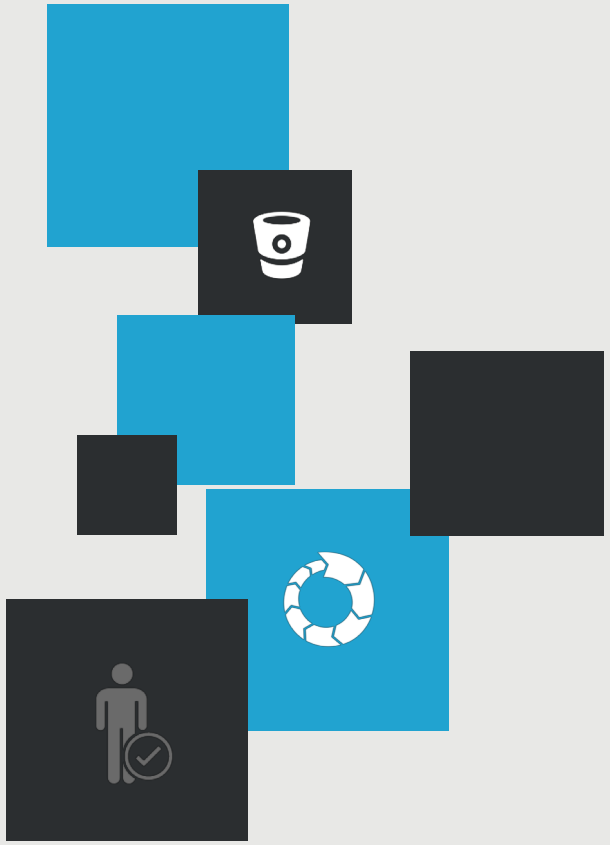
Contents

Contents

1

TEAM INTRO

TEAM



Zekun Tang

Graduate student of Lanzhou University
Kaggle master
Several top 3 in the Alibaba Tianchi competition

Qi Wu

intern at Microsoft Research
TJU SmartSafteyResearchGroup
Kaggle Master

WEILONG CHEN

Tencent Wechat Intern
UESTC IntelliGame lab
WSDMCUP 2019/2020 **TOP 1**
ACM MM Grand Challenge 2019/2020 **TOP 1/TOP 3**

Wei Bao

Graduate student in Southeast University
WSDM2020 TOP1
SemEval2020 task3 TOP1

SIGIR 2020
XI'AN·CHINA

Contents

2

BACKGROUND

BACKGROUND

How to predict the price trend of bulk commodities?

This competition focuses on 6 non-ferrous metals, predicting the direction (up/down) of non-ferrous metal prices in three time periods (**1day 20day 60day**).

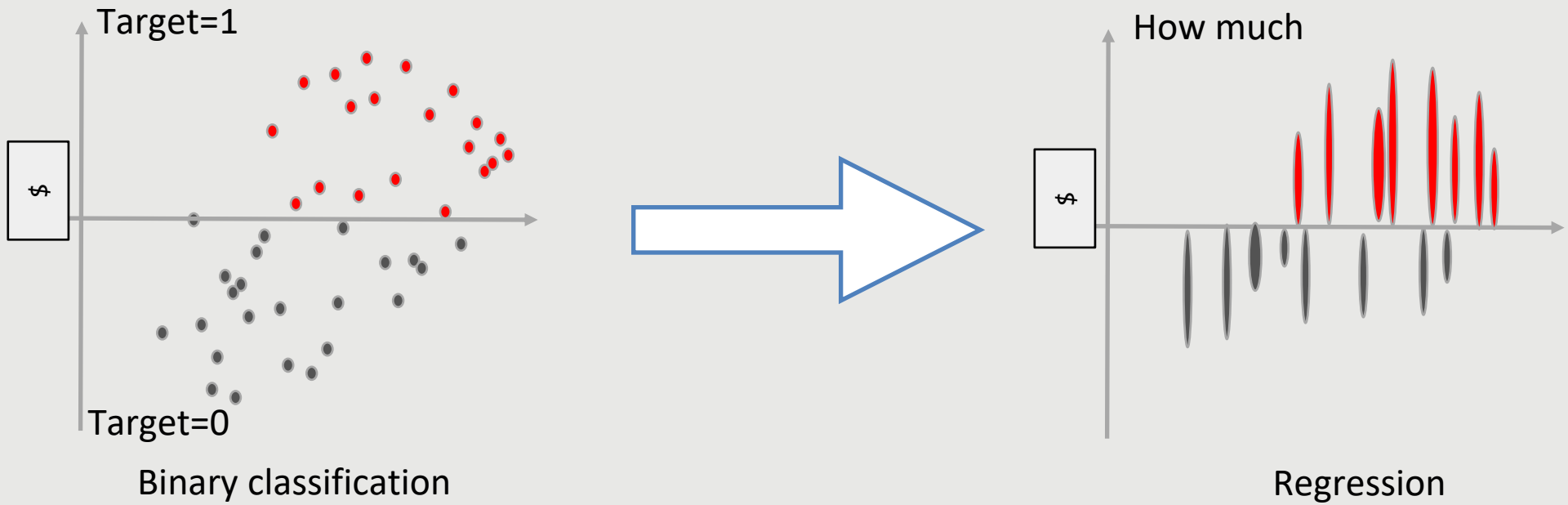
LMEAluminium_OI: Aluminium open interest
LMECopper_OI: Copper open interest
LMENickel_OI: Nickel open interest
LMELead_OI: Lead open interest
LMETin_OI: Tin open interest
LMEZinc_OI: Zinc open interest
LMECopper3M_longer: Copper 3 months data
LMELead3M: Lead 3 months data
LMENickle3M: Nickle 3 months data
LMETin3M: Tin 3 months data
LMEZinc3M: Zinc 3 months data
LMEAluminium3M: Aluminium 3 months data

Contents

3

METHODOLOGY

METHODOLOGY

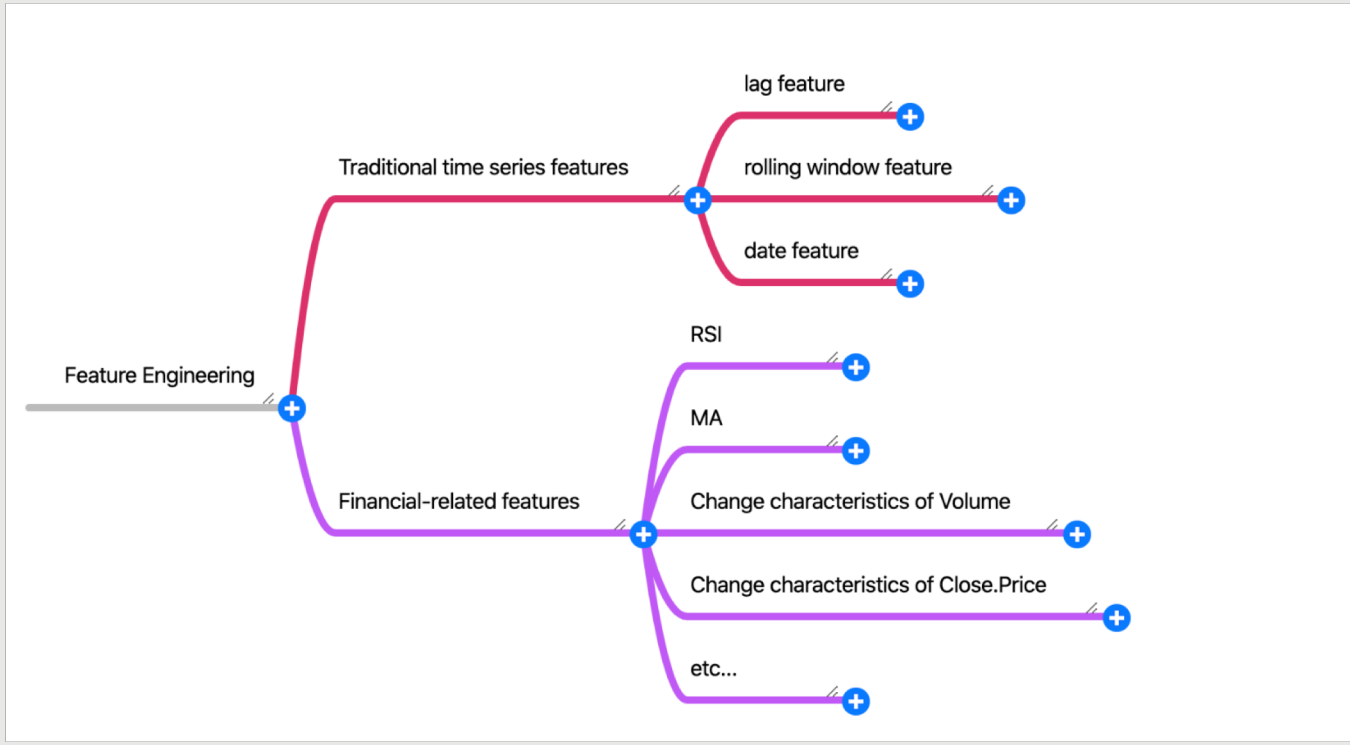


Binary classification

Regression

- The goal is a **binary task**.
- However, we believe that in the case of fluctuations, the two-category modeling cannot well reflect the strength (range) of the fluctuations.
- Therefore, we convert its target to use **regression**.

METHODOLOGY——FEATURE ENGINIEERING



caption

Our characteristics are mainly divided into two parts:

- **Traditional time series features**
- **Some financial-related features**

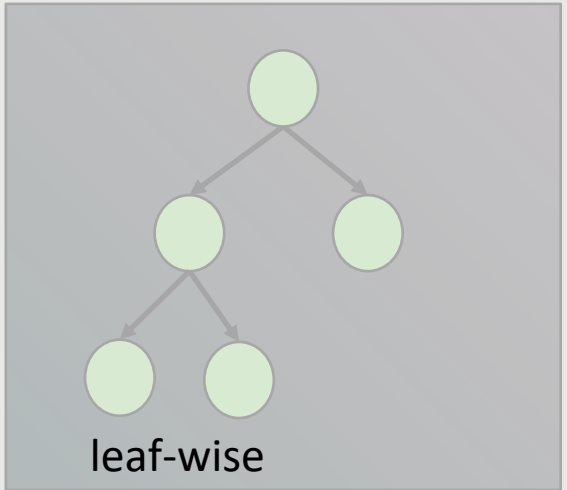
- **Adopted different window and lag values**

for different targets t1, t20, and t60.

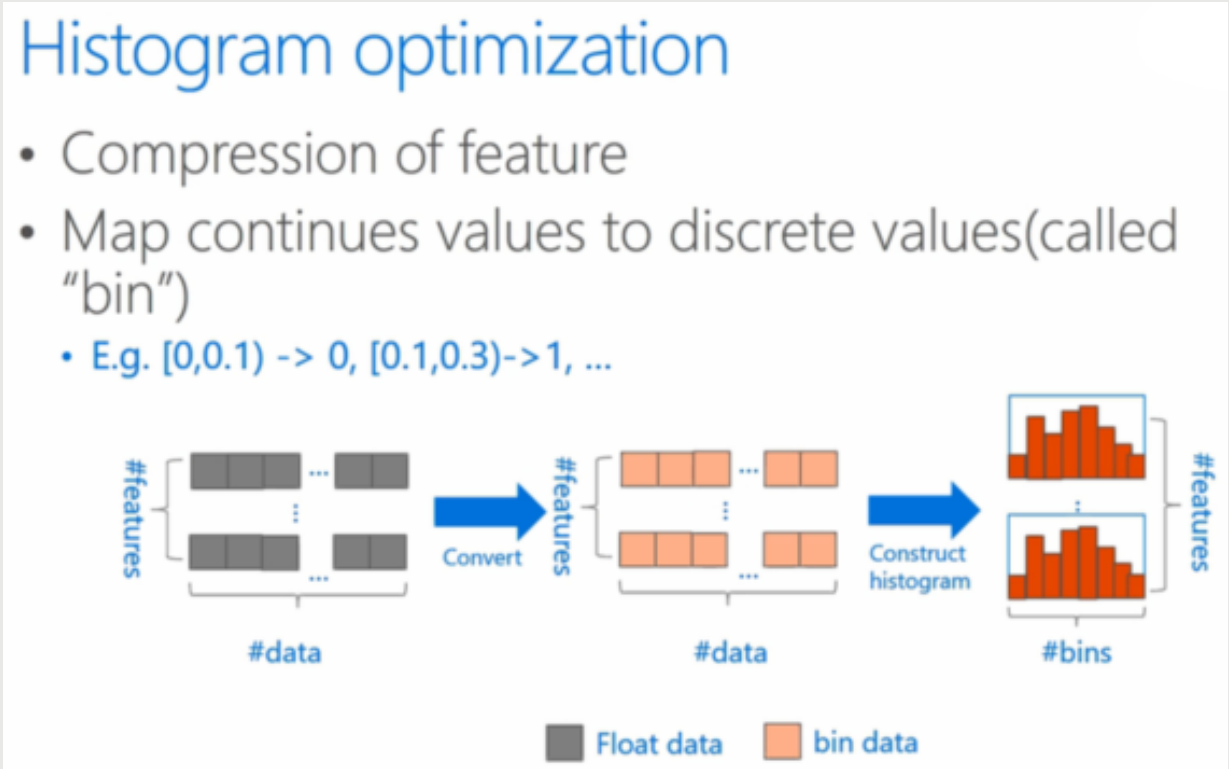
METHODOLOGY——MODEL

The model we chose for this competition is **LightGBM**

- learning rate: 0.003
- sub_row parameter: 0.1-0.9
- num_leaves: 15-63
- min_data_in_leaf: 40-60

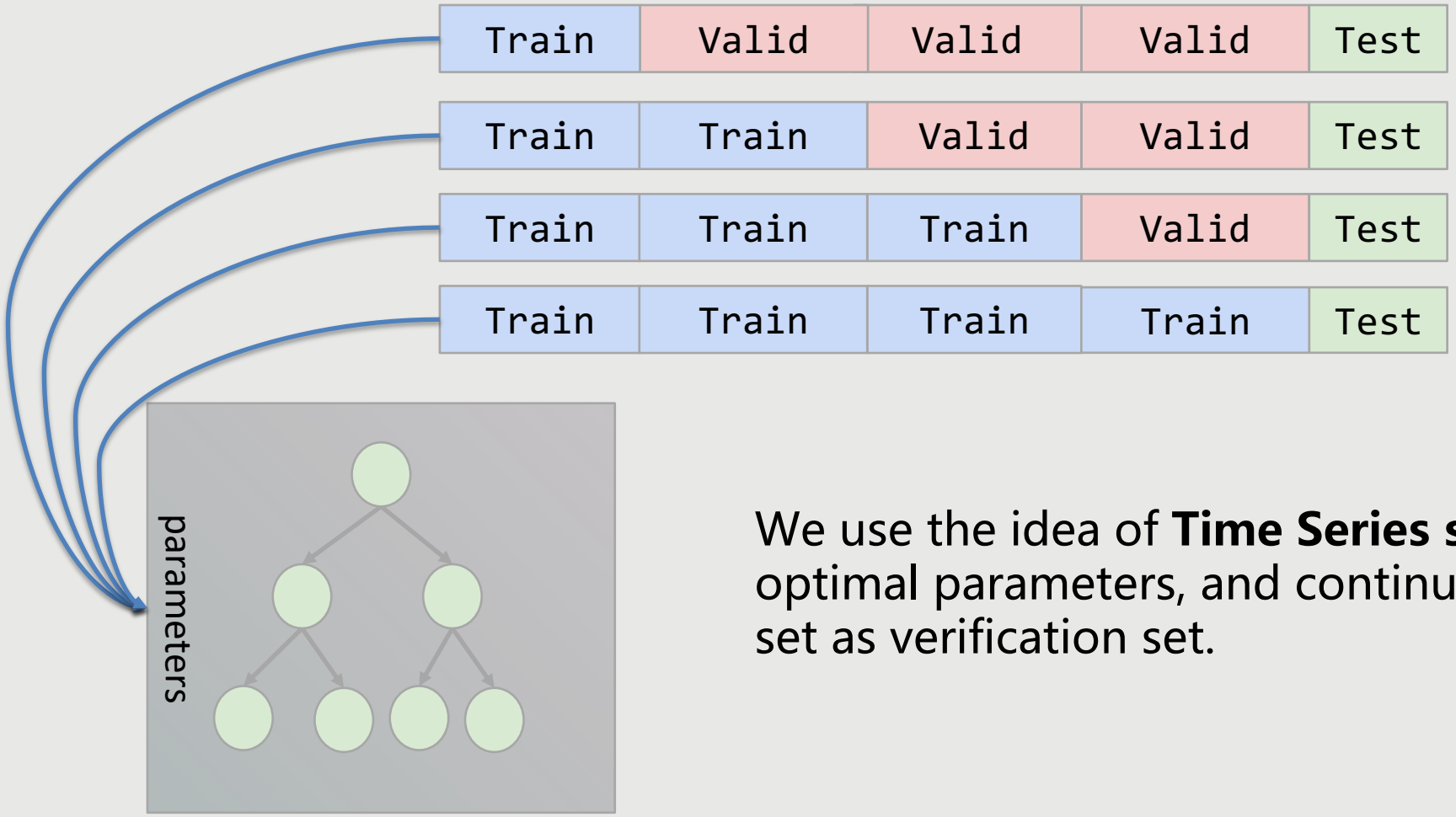


leaf-wise



caption

METHODOLOGY——VALID METHOD



We use the idea of **Time Series split** to select the optimal parameters, and continue to use the latest data set as verification set.

Contents

4

FUTURE WORK

FUTURE WORK

- **Professional feature mining is still not enough**
- **The model has not been well adjusted**
- **The potential of the deep model has not yet been realized**

Thank you!


1909
LANZHOU UNIVERSITY


天津大學
Tianjin University


UESTC
1956


SOUTHEAST UNIVERSITY
東南大學
1902
南京


SIGIR 2020
XI'AN·CHINA