Final Project Progress Check

Group 11

👑Debby(King of the IDEAL Lab, Boss of the Kingdom, $300M employer)

👸Kevin(Singer of the Island, Wanderer of the World)

😏Jerry(Personal bodyguard, UFC lightweight champion)

🚀Brian(Dark knight, Revenger from the corruption)

🈹Louis(Dishwasher, IDEAL lab Owner and Mentor)

Rank Result

Using XGboost   dataset: cust_info

#	隊伍名稱	成員	提交次數	分數	上傳時間
71	理想實驗室	5	4	0.015129	12/4/2022 12:13:37 AM

Using LSTM   dataset: cust_info, dp

#	隊伍名稱	成員	提交次數	分數	上傳時間
45	理想實驗室	5	7	0.022779	12/11/2022 11:36:44 PM

Current Preprocessing

Main Principles

	Categorical	Numerical
01 Missing Value(fill in)	Mode Value	Median Value
02 Outlier		Set Upper/Lower bound (using boxplot)
03 Normalize		min/max scale to (0,1)
04 Transform	One-hot encode	

Current Preprocessing

Dataset Preparation

Customer Info

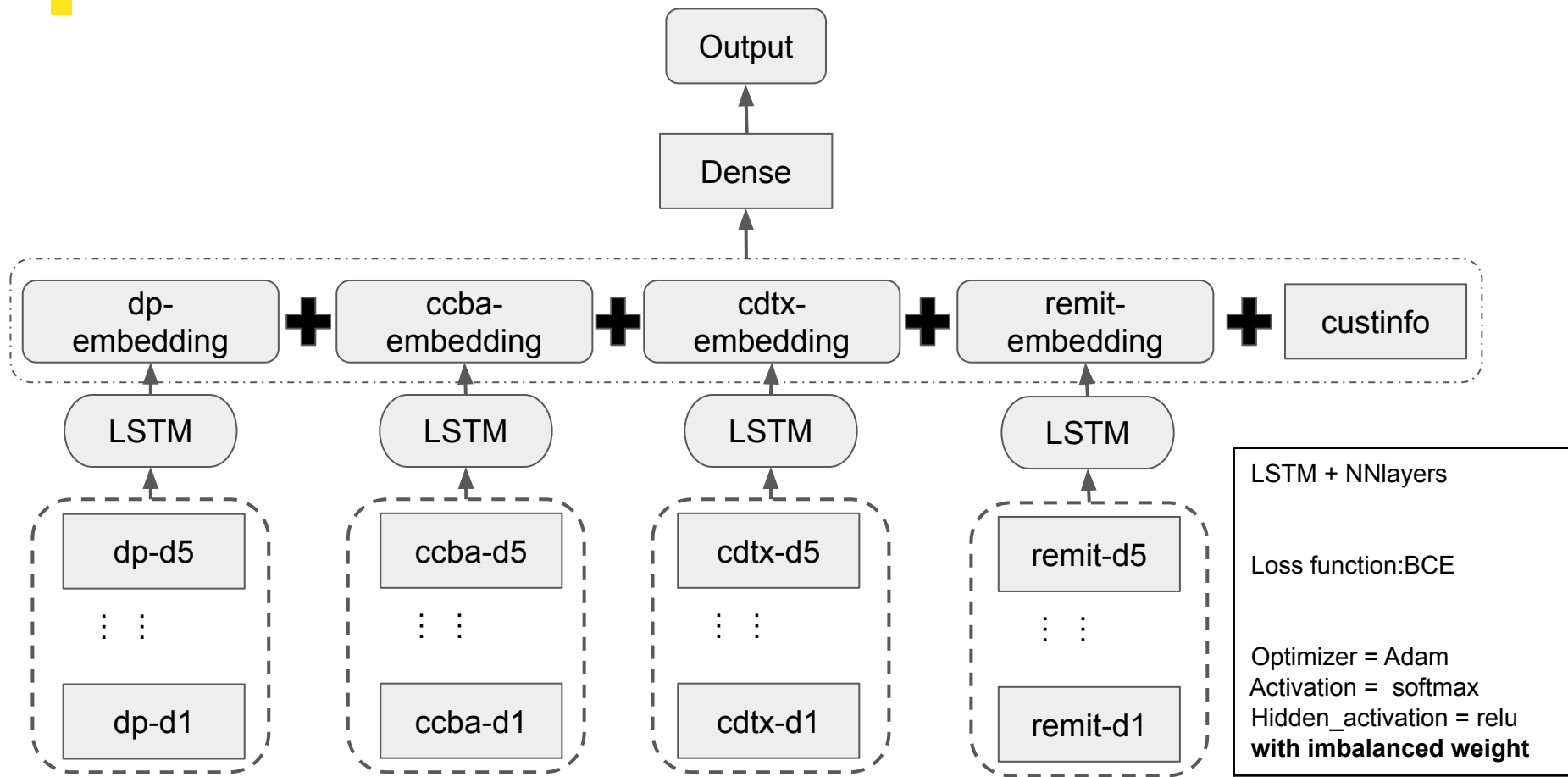|		
cust_info_ count	occurrence	numeric
dp_count	occurrence	numeric
ccba_ count	occurrence	numeric
cdtx_count	occurrence	numeric
remit_ count	occurrence	numeric

dp

total_amount	change to NTD currency	numeric
foreign	Whether international exchage	category(0/1)
cash_from_ counter	Whether change from counter	category(0/1)

cdtx

is_in _taiwan	whether transact in Taiwan	cat(0/1)
is_ taiwan_ cur	whether transact in NTD	cat(0/1)

Model Structure

Future Work

- Optimize the way of filling missing value.
- Finish the remaining dataset preprocessing and include them in our model training structure.
- Try out different model structure to improve the result.
- More data exploration to find out or generate more meaningful feature.