

# Fall 2023 MGMT 571 Final Project

Deadline: 11/29/2023 6:00pm EST

This project aims to apply data mining algorithms covered in the course in a real data competition. If you are new to Kaggle, please make sure to read all instructions on this link of data competition: <https://www.kaggle.com/docs/competitions>. You need to first register a Kaggle account before any submission.

1. We would like to develop a predictive model that combines various econometric measures to predict if a firm will file for bankruptcy. The details about our final project competition, data and evaluation criterion are available in the following website. **Do not share this link with others!**

<https://www.kaggle.com/t/01d3a13d8db44fee8a9bab10e66dcf43>

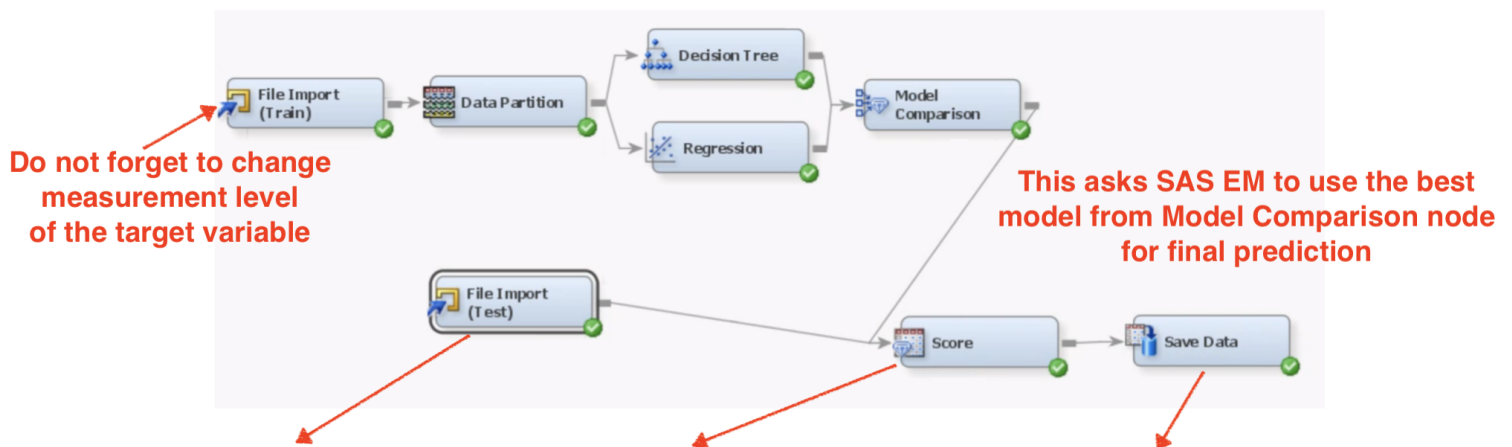
2. The final project contributes 20 points to your final score: (1) 10 points will be evaluated based on the rank on the **Private leaderboard** (Not Public leaderboard!), see below the grading policy; (2) 10 points will be based on presentation.

Top 0% ~ 10% :	10/10;
Top 10% ~ 30% :	9/10;
Top 30% ~ 50% :	8/10;
Top 50% ~ 70% :	7/10;
Top 70% ~ 90% :	6/10;
Top 90% ~ 100% :	5/10.

3. Rules in Kaggle competition: (1) Only use your Team Name to submit solutions, see the third column in Fall2023\_MGMT571\_teams.xlsx; (2) Maximal daily submission is 10; (3) The number of submissions eligible for the final private leaderboard is 2; (4) Only use SAS EM in your analysis. **For the fairness of the whole class, using anything other than SAS EM for your competition and analysis will cause your team to receive ZERO on the Final Project;** (5) Only use the provided training data to train your model. **If any of this rule is broken, the final score will be discounted.**
4. Please choose one team member to submit a brief description of your team's final chosen two models for Private leaderboard and **a zip file of your SAS EM project** to Brightspace, **before the project deadline**, so that the TA can reproduce your result on Private leaderboard. **If TA can not reproduce your result, your final score will be discounted.**
5. The presentation will be evaluated based on the content and presentation skills. The presentation should be  $\leq$  5-minutes long. The content should contain the final algorithm used in the leaderboard, other tried algorithms, why do you choose the final one, lessons learnt etc. The presentation will be during the class time on 12/4/2023. The order of presentation is the order shown in Fall2023\_MGMT571\_teams.xlsx

# Tips on Final Project

Given a training data and a test dat with only predictors, how can we make a prediction and save the prediction to a csv file?



Property	Value
<b>General</b>	
Node ID	FIMPORT2
Imported Data	
Exported Data	
Notes	
<b>Train</b>	
Variables	
Import File	W:\MGMT571\Data\bankrupt...
Maximum rows to import	1000000
Maximum columns to import	10000
Delimiter	,
Name Row	Yes
Number of rows to skip	0
Guessing Rows	500
File Location	Local
File Type	csv
Advanced Advisor	No
Rerun	No
<b>Score</b>	
Role	Score
<b>Report</b>	
Summarize	No

Property	Value
<b>General</b>	
Node ID	Score
Imported Data	
Exported Data	
Notes	
<b>Train</b>	
Variables	
Type of Scored Data	Data
Use Fixed Output Names	Yes
Hide Variables	No
Hide Selection	
<b>Score Data</b>	
Validation	No
Test	No
<b>Score Code Generation</b>	
Optimized Code	Yes
C Score	No
Java Score	No
Java Package Name	Default
User Package Name	
<b>Report</b>	
Graphical Reports	Yes

Property	Value
<b>General</b>	
Node ID	EMSave
Imported Data	
Exported Data	
Notes	
<b>Train</b>	
<b>Output Options</b>	
Variables	
Filename Prefix	bankruptcy_myforecast
Replace Existing Files	Yes
All Observations	Yes
Number of Observations	1000
<b>Output Format</b>	
File Format	Comma-separated Values (.csv)
SAS Library Name	
Directory	W:\MGMT571\Data
<b>Output Data</b>	
All Roles	Yes folder for saving output
Select Roles	

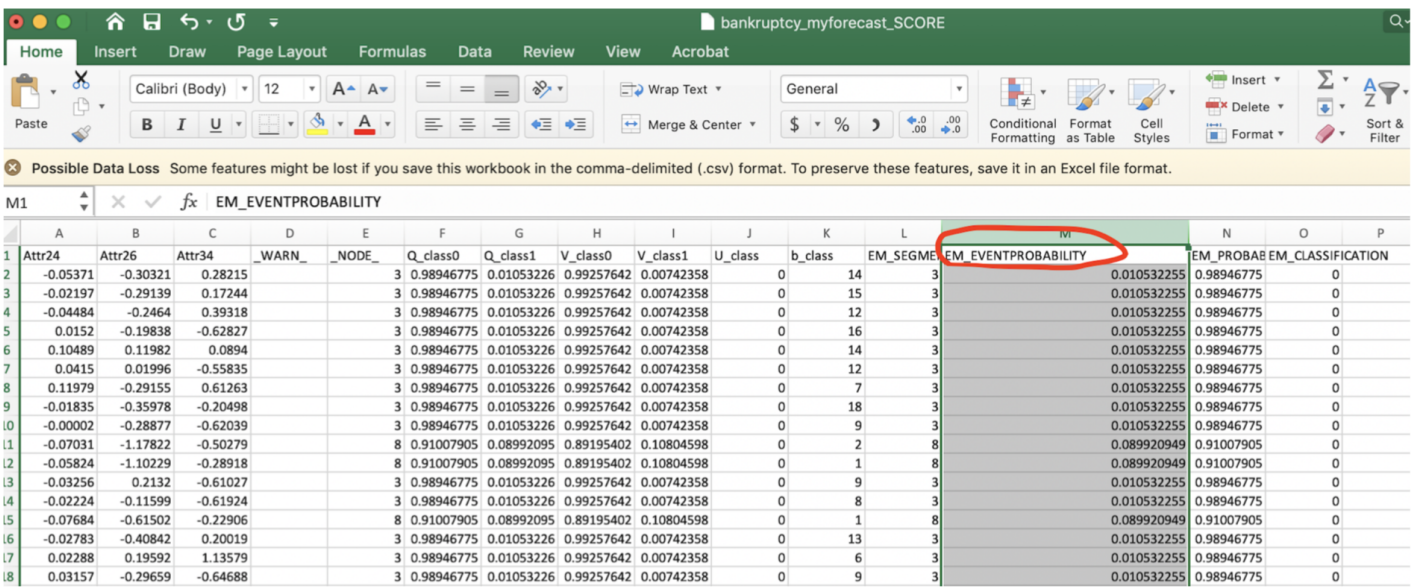
## Tips on Final Project

Then we go to the folder where the data is saved to. Open the one with file name "X\_SCORE.csv", where X is the filename prefix you specified in Save Data node.

Filename	Size
Advertising.csv	4.1 KB
<b>bankruptcy_myforecast_SCORE.csv</b>	<b>533.7 KB</b>
bankruptcy_myforecast_TRAIN.csv	983.5 KB
bankruptcy_myforecast_VALIDATE.csv	422.9 KB
bankruptcy_sample_submission.csv	58.3 KB
bankruptcy_Test_X.csv	2.8 MB
bankruptcy_Train.csv	5.5 MB
Beef.csv	774 B
beef.sas7bdat	131.1 KB
catalog.csv	135.1 KB
Churn.csv	303.9 KB
Churn.sas7bdat	655.4 KB
Credit.numbers	326.2 KB
Default.csv	306.2 KB
Hitters.csv	17.3 KB
HMEQ.csv	403.2 KB
hmeq.sas7bdat	720.9 KB
Pharmacy_test.csv	2.1 KB
Pharmacy_train.csv	9.6 KB
Pharmacy.csv	13.1 KB

# Tips on Final Project

Copy values of "EM\_EVENTPROBABILITY", which refers to estimated probability of bankruptcy for test data, to the second column of bankruptcy\_sample\_submission.csv. Make sure column names of csv file are still "ID" and "class". Submit it to Kaggle.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
	Attr24	Attr26	Attr34	_WARN_	_NODE_	Q_class0	Q_class1	V_class0	V_class1	U_class	b_class	EM_SEGME	EM_EVENTPROBABILITY	EM_PROBAE	EM_CLASSIFICATION	
1	-0.05371	-0.30321	0.28215			3	0.98946775	0.01053226	0.99257642	0.00742358	0	14	3	0.010532255	0.98946775	0
2	-0.02197	-0.29139	0.17244			3	0.98946775	0.01053226	0.99257642	0.00742358	0	15	3	0.010532255	0.98946775	0
3	-0.04484	-0.2464	0.39318			3	0.98946775	0.01053226	0.99257642	0.00742358	0	12	3	0.010532255	0.98946775	0
4	0.0152	-0.19838	-0.62827			3	0.98946775	0.01053226	0.99257642	0.00742358	0	16	3	0.010532255	0.98946775	0
5	0.10489	0.11982	0.0894			3	0.98946775	0.01053226	0.99257642	0.00742358	0	14	3	0.010532255	0.98946775	0
6	0.0415	0.01996	-0.55835			3	0.98946775	0.01053226	0.99257642	0.00742358	0	12	3	0.010532255	0.98946775	0
7	0.11979	-0.29155	0.61263			3	0.98946775	0.01053226	0.99257642	0.00742358	0	7	3	0.010532255	0.98946775	0
8	-0.01835	-0.35978	-0.20498			3	0.98946775	0.01053226	0.99257642	0.00742358	0	18	3	0.010532255	0.98946775	0
9	-0.00002	-0.28877	-0.62039			3	0.98946775	0.01053226	0.99257642	0.00742358	0	9	3	0.010532255	0.98946775	0
10	-0.07031	-1.17822	-0.50279			8	0.91007905	0.08992095	0.89195402	0.10804598	0	2	8	0.089920949	0.91007905	0
11	-0.05824	-1.10229	-0.28918			8	0.91007905	0.08992095	0.89195402	0.10804598	0	1	8	0.089920949	0.91007905	0
12	-0.03256	0.2132	-0.61027			3	0.98946775	0.01053226	0.99257642	0.00742358	0	9	3	0.010532255	0.98946775	0
13	-0.02224	-0.11599	-0.61924			3	0.98946775	0.01053226	0.99257642	0.00742358	0	8	3	0.010532255	0.98946775	0
14	-0.07684	-0.61502	-0.22906			8	0.91007905	0.08992095	0.89195402	0.10804598	0	1	8	0.089920949	0.91007905	0
15	-0.02783	-0.40842	0.20019			3	0.98946775	0.01053226	0.99257642	0.00742358	0	13	3	0.010532255	0.98946775	0
16	0.02288	0.19592	1.13579			3	0.98946775	0.01053226	0.99257642	0.00742358	0	6	3	0.010532255	0.98946775	0
17	0.03157	-0.29659	-0.64688			3	0.98946775	0.01053226	0.99257642	0.00742358	0	9	3	0.010532255	0.98946775	0