INFSCI 2595
Machine Learning

Final project

Test Set

Submission report for bonus

A hold out test set is provided on Canvas.
The x-variables and v-variables are provided in separate CSV files

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	x01	x02	x03	x04	x05	x06	x07	x08	x09	x10	x11	x12	x13	x14	x15	x16	x17	x18
2	0.94073	0.05828	0.00503	0.59886	0.98784	0.23289	0.45587	0.32289	0.40582	0.55326	0.08012	0.30013	0.17891	0.1823	0.86184	0.2867	0.70158	0.7
3	0.76027	0.94135	0.07989	0.92071	0.40691	0.02292	0.30672	0.28628	0.96045	0.9737	0.00837	0.72043	0.4003	0.9472	0.37346	0.09457	0.95002	0.7
4	0.64633	0.11665	0.23591	0.30762	0.81988	0.60267	0.35525	0.23612	0.02303	0.00663	0.93939	0.45451	0.26105	0.03169	0.33784	0.72057	0.98321	0.7
5	0.9584	0.82805	0.86061	0.36849	0.16157	0.09947	0.37973	0.31822	0.73498	0.01769	0.09886	0.00279	0.96991	0.30515	0.90831	0.75641	0.66405	0.7
6	0.08984	0.20511	0.98015	0.48611	0.93347	0.0445	0.52616	0.22381	0.70407	0.97177	0.17135	0.73531	0.76489	0.02836	0.64684	0.78115	0.47577	0.7
7	0.72444	0.0403	0.02067	0.01861	0.06261	0.87538	0.32816	0.24967	0.86593	0.97968	0.64035	0.03184	0.42726	0.46555	0.78404	0.93761	0.26548	0.7
8	0.19842	0.68885	0.97443	0.97707	0.88912	0.77723	0.22392	0.32392	0.72005	0.35781	0.87367	0.95098	0.31942	0.12309	0.88176	0.93906	0.25162	0.7
9	0.8102	0.04044	0.08934	0.8121	0.14597	0.24745	0.41825	0.1438	0.30591	0.98965	0.80907	0.20438	0.27845	0.29912	0.00172	0.98056	0.93669	0.
10	0.14752	0.39208	0.21318	0.91892	0.65103	0.91303	0.59282	0.18486	0.84061	0.03895	0.67334	0.95197	0.08784	0.41394	0.39273	0.15254	0.01093	0.7
11	0.87882	0.65899	0.74753	0.54536	0.30324	0.84349	0.32243	0.20598	0.10785	0.03082	0.06876	0.93541	0.21324	0.78957	0.39834	0.21814	0.02955	0.7

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	v01	v02	v03	v04	v05	v06	v07	v08	v09	v10	v11	v12	v13	v14	v15	v16	v17	v18
2	0.405817	0.496902	0.624771	0.356669	0.531445	0.129793	0.573198	0.133672	0.582616	0.115048	0.553257	0.115048	0.484534	0.023702	0.41168	0.019839	0.335983	0.020
3	0.960454	0.664385	0.78576	0.38519	0.710199	0.248861	0.625308	0.220654	0.685504	0.195037	0.973696	0.195037	0.643591	0.196808	0.72052	0.177527	0.750481	0.194
4	0.023031	0.132775	0.146187	0.144678	0.110786	0.257348	0.085005	0.255552	0.074645	0.224594	0.006632	0.224594	0.243843	0.63934	0.208862	0.610472	0.170633	0.879
5	0.734977	0.407813	0.225047	0.243572	0.196151	0.233206	0.028364	0.224582	0.02303	0.208853	0.017688	0.208853	0.188366	0.284986	0.169217	0.46627	0.122564	0.598
6	0.704066	0.734892	0.205695	0.462744	0.379698	0.270848	0.904615	0.29124	0.920627	0.247531	0.971774	0.247531	0.438949	0.075966	0.316485	0.079102	0.204139	0.048
7	0.865929	0.427631	0.75844	0.535996	0.681616	0.690655	1	0.677146	0.980878	0.643557	0.979684	0.643557	0.629861	0.018889	0.648188	0.035232	0.674863	0.047
8	0.720052	0.728545	0.13768	0.694901	0.201377	0.749917	0.271069	0.720206	0.295388	0.711179	0.357814	0.711179	0.548256	0.2199	0.418947	0.179044	0.360885	0.182
9	0.305908	0.172454	0.839588	0.128021	0.755124	0.432664	0.954355	0.437061	0.990042	0.386541	0.989647	0.386541	0.411674	0.157572	0.345211	0.168993	0.279906	0.261
10	0.840614	0.710198	0.041097	0.748201	0.032286	0.664163	0.075101	0.707735	0.078623	0.652764	0.038951	0.652764	0.60315	0.124464	0.555721	0.085089	0.530807	0.081

The input names are the same as those in the training set. ALL x-variables and v-variables are present in the CSV files. Only a few of them are shown here.

You must predict the continuous response and the binary outcome using this test set

- You must select 1 regression model and 1 classification model.

- You must predict the continuous output.

- You must predict the probability of the event.

- You must classify the binary outcome assuming a default threshold of 0.5.

Organize the test set predictions

- Compile the predictions into a dataframe with 4 columns:
  - `id` – the row index (use `tibble::rowid_to_column()` function)
  - `response` – the prediction for the continuous response
  - `outcome` – the classified outcome
    - Must have values `event` and `non_event`
  - `probability` – the predicted probability of the event

- Save the dataframe to a CSV file.
  - Can save using the `readr::write_csv()` function

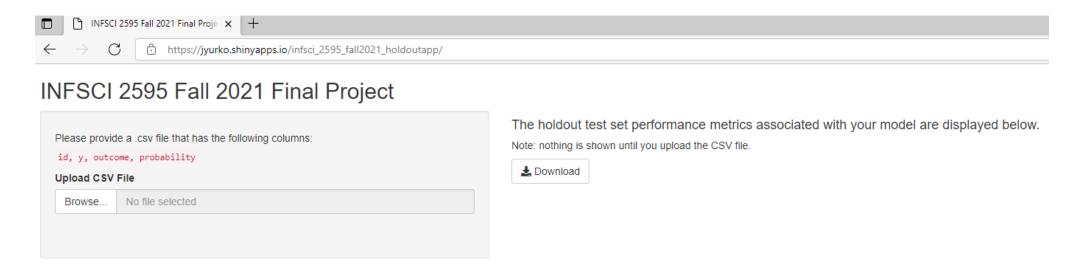# Canvas includes example Markdowns which show how to the compile the predictions

- The example Markdowns train a few bad models, show how to make predictions, and how to compile those predictions accordingly.

- Two Markdowns are provided on Canvas.

  - tidymodels example
  - caret example

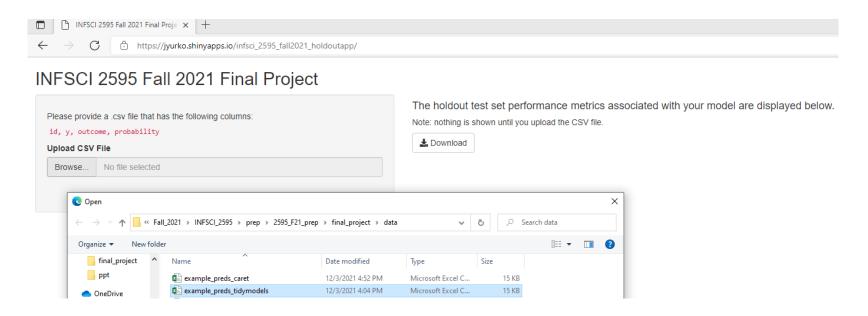Snippet of the compiled predictions and saved into a CSV file

	A	B	C	D
1	id	response	outcome	probability
2	1	0.786915	non_even	0.335429
3	2	0.87442	non_even	0.328427
4	3	-0.64897	event	0.733724
5	4	0.707465	non_even	0.362065
6	5	0.559978	non_even	0.441181
7	6	-0.15269	event	0.597416
8	7	-0.61458	event	0.737261
9	8	-0.42231	event	0.666687
10	9	-0.2149	event	0.639151
11	10	0.749017	non_even	0.357068

Your predictions will "scored" by uploading the predictions to a website

- Go to the following [R Shiny app](#).

- The landing page looks like:

Your predictions will "scored" by uploading the predictions to a website
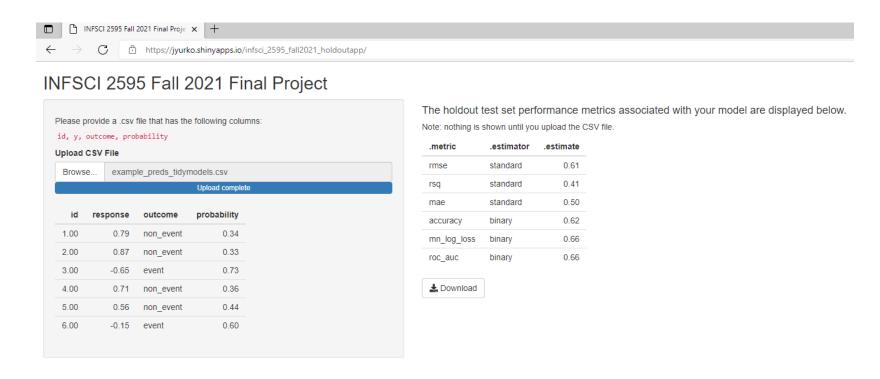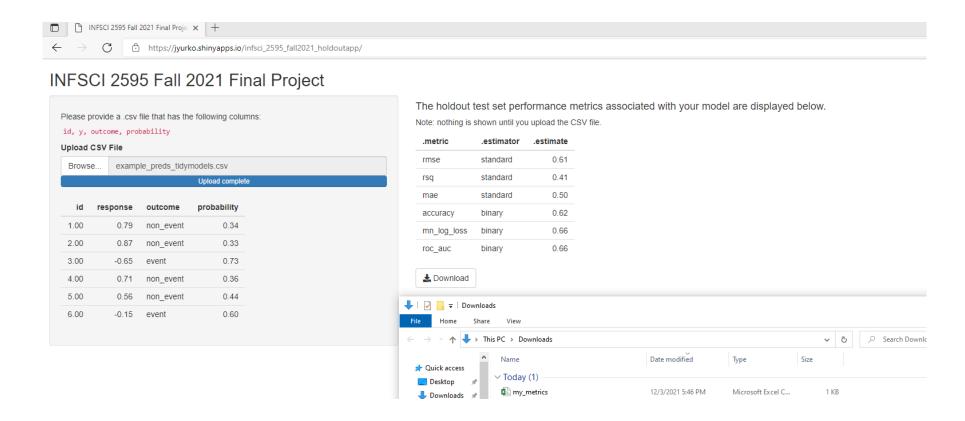
- Go to the following [R Shiny app](#).
- Select the Browse button and upload your CSV file of predictions to the website.
- I named my example CSV file example_preds_tidymodels.csv
- You may name your CSV file whatever you want.

Your predictions will "scored" by uploading the predictions to a website

- Once uploaded the performance metrics on the hold-out test set will be shown to you.

Your predictions will "scored" by uploading the predictions to a website

- Press the Download button to save the performance metrics to your computer. The downloaded CSV file is named my_metrics.csv

You **MUST** submit the downloaded CSV file as part of your final project submission

- The downloaded CSV file must be uploaded to Canvas along with all of your rendered HTML files and source .Rmd files.

- Do **NOT** zip files!!!!!!!!!!!! Upload each file.

BONUS: 10 points

- Create a short presentation which shows:
  - What are the most important inputs?
  - What are the trends of the continuous response with respect to the most important inputs?
  - What are the trends of the event probability with respect to the most important inputs?
  - What input values do you recommend to minimize the continuous response?
  - What input values do you recommend to minimize the event probability

- Submit your presentation as a Power Point .pptx file.