Guideline: Prepare Final Submission Dataset

* Note, this Guideline has been updated in March 2023.

This is to provide you with some guideline on the preparation of the scored dataset for final submission (with 3 columns) as required by the project description. Read the below steps and explanations carefully, and assuming you have the data objects that I describe below, you should be able to follow the steps exactly to get the final submission dataset (CSV with 3 columns).

Assume you have read in the original SCORE data as a dataframe `df_Score`, and it has the following column labels:

```
['id_new', 'age', 'gender', 'tenure', 'region', 'dl', 'has_v_insurance',
'v_age', 'v_accident', 'v_prem_quote', 'cs_rep']
```

Follow the guideline provided in the course materials on how to properly preprocess this SCORE data. (hint: the preprocessing of SCORE data is very similar to that of the TEST data but quite different from that of the TRAIN data, assuming you use what we refer to as `transformers`). Afterwards, be sure to convert the data into `NumPy` array.

Important! Assume that your model will produce a 2000 x 2 `NumPy` array as the activated values using the preprocessed input data of SCORE (must be a `NumPy` array), as in:

```
a_out_score.shape
```
output: (2000, 2)

```
isinstance(a_out_score, np.ndarray)
```
output: True

Then follow these steps after you have obtained `a_out_score`:

```
# Make the final submission dataset based on df_Score
df_Submit = df_Score[['id_new']].copy()
```

```
# Add the predicted probabilities (only one column, the probability for label 1 only) to this dataset
df_Submit['probability'] = a_out_score[:, 1]
```

```
# sanity check: print a few rows
display(df_Submit.head())
```

```
# Add the classification column
df_Submit['classification'] = np.argmax(a_out_score, axis = -1)

# sanity check: print a few rows
display(df_Submit.head())

# export dataset
df_Submit.to_csv('my_prediction.csv', index=False)
```

You will find a new csv file created in the same directory where your Jupyter notebook is located.