Crash Kaggle Competition Tutorial

> for STAT3009: Recommender Systems

Participating in a Kaggle competition can be a rewarding experience that helps you improve your data science skills. Here's a step-by-step tutorial to guide you through the process, along with examples.

Create a Kaggle Account

1. **Sign Up**: Go to Kaggle and create an account if you don't already have one.
2. **Explore Competitions**: Navigate to the "Competitions" tab to see ongoing and upcoming competitions.

Choose a Competition

1. **Select a Competition**: Pick a competition that interests you. For example, the "Titanic: Machine Learning from Disaster" competition.
2. **Read the Overview**: Understand the problem statement, evaluation metric, and data description.

> Important
>
> Here you have two methods to submit your solution:
>
> - Submit a Kaggle notebook (see this section)
> - Run the code in Google Colab (or your local machine) and upload prediction to Kaggle (see this section)

Run the Kaggle Notebook

1. **Create a New Notebook**:
    - Navigate to the "Code" tab in the competition page.
    - Click on "New Notebook" to create a Kaggle Jupyter Notebook.

2. **Load the Data**:

```
# Load the training data directly from Kaggle datasets
train_data = pd.read_csv('/kaggle/input/<your-competition-name>/train.csv')
```

3. **Data Exploration and Processing**: Use the notebook to explore and preprocess your data as needed.

4. **Make Predictions**: Run your prediction code directly in the notebook.

5. **Prepare Submission**:
   In your notebook, create a DataFrame for your submission and save it as `submission.csv`.

```
# Load the sample submission file to understand the required format
sample_submission = pd.read_csv('/kaggle/input/<your-competition-name>/sample_submission.csv
submission = sample_submission.copy()

# Create a DataFrame for your submission
submission['<the-target-col-name>'] = <your-predictions>

# Save to CSV
submission.to_csv('submission.csv', index=False)
```

6. **Submit Your Notebook**:
   - After running your notebook and generating predictions, click on the "`Save Version`" button at the top right of the notebook interface.
   - Now your notebook is ready and saved, so you can go to the Notebook Viewer and jump to the "`Output`" section by selecting it from the right side of the page as shown in the figure below.
   - Here you will find a "`Submit`" button in blue color, whenever you are ready to submit your notebook, click on that button as shown in the figure below.
   - See this link for more details.

Run the Code Colab or Locally

1. **Download the Data**: Navigate to the "Data" tab of the competition page and download the datasets (e.g., `train.csv`, `test.csv`).

2. **Set Up Your Environment**:
   - For Google Colab, you can start a new notebook and upload your datasets or load them directly from Kaggle.
   - If using a local machine, ensure you have Python and necessary libraries (like pandas, numpy, and matplotlib) installed.

3. **Load and Explore the Data**:

```
import pandas as pd

# Load the training data
```

```
train_data = pd.read_csv('train.csv')
print(train_data.head())
```

4. **Make Predictions**: After processing and analyzing the data, generate predictions based on your model for `test.csv`.

5. **Prepare Submission**:

```
# Load the sample submission file to understand the required format
sample_submission = pd.read_csv('sample_submission.csv')
submission = sample_submission.copy()

# Create a DataFrame for your submission
submission['<the-target-col-name>'] = <your-predictions>

# Save to CSV
submission.to_csv('submission.csv', index=False)
```

6. **Submit to Kaggle**:
    - Go to the "Submit Predictions" tab on the competition page.
    - Upload your `submission.csv` file and click "Submit".

Conclusion

Participating in Kaggle competitions involves data exploration and understanding how to submit your predictions effectively. You can either download datasets and submit a CSV file or use Kaggle's Jupyter Notebook environment to make and submit predictions directly. Happy Kaggle-ing!