Making your first submission on Numerai notebook is usually first attempt of newcomer like me, but with latest V4 “supermassive” dataset it will crash in google Colab due to 12GB RAM limitation.

You can fix this by downcasting features of read dataframes e.g. like this:

download the latest training dataset (takes around 30s) - original code

training_data = pd.read_csv("https://numerai-public-datasets.s3-us-west-2.amazonaws.com/latest_numerai_training_data.csv.xz") training_data.head()

added code to downcast feature columns

find only the feature columns

feature_cols = training_data.columns[training_data.columns.str.startswith('feature')]

conserve memory by converting feature data to np.float16

training_data[feature_cols] = training_data[feature_cols].astype(np.float16)

similarly after reading tournament data:

download the latest tournament dataset - original code (now takes around 2 mins)

tournament_data = pd.read_csv("https://numerai-public-datasets.s3-us-west-2.amazonaws.com/latest_numerai_tournament_data.csv.xz") tournament_data.head()

added code to downcast feature columns

tournament_data[feature_cols] = tournament_data[feature_cols].astype(np.float16)

Alternatively you can use kaggle notebooks which have RAM limit at 16GB, but even this is will not suffice without downcasting or using latest *_int8.parquet

datasets. I have copied First submission notebook to kaggle, downcasted data to fit available RAM and made it public, so that newcomers can finish their first attempt.