

Part (A)

The 2006 Netflix prize was an open call to improve the accuracy of the predictions of Netflix viewers' ratings of films. Sponsored by Netflix, the winning team with the algorithm that has the greatest improvement in root mean squared error (RMSE) over Netflix's internal algorithm, CineMatch, would win a cash prize. The equation of RMSE is seen below, with \hat{y} as the predicted rating and y as the actual rating in the dataset.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$

Although all teams who submitted their algorithms were judged, the threshold to qualify for the grand cash prize of \$1 million, is obtaining a 10% improvement in accuracy in predicting 3 million ratings, compared to CineMatch. If no one passes this threshold, Netflix will award \$50,000 progress prize to the team that makes the most significant advancement towards obtaining a 10% improvement in accuracy.

At the beginning of the contest, the most commonly used method was nearest neighbors. This collaborative filtering method predicts these ratings by weighing the average rating of similar items by the same user, using Pearson correlation, cosine similarity etc. Bell et al. describe three problems with this method: firstly, the choice of similarity is random, highly correlated neighbors are "double counted" and the predictions are not as reliable for movies with few neighbors (page 26).

The best models in the Netflix prize were combined sets of predictions from multiple models. The winning model was a "linear combination of 107 prediction sets" with weights determined by ridge regression" (page 29). One characteristic that improved the overall prediction was if the blend of these algorithms was not highly correlated with the other components (page 28).

Part (B)

Project euler friend key:

1407611_M6JGHRfytmvLFQ3C6vW9OztwQHGAOdQP

Project euler username:

joesetan

The answer for question 1 is 233168.

The code to obtain this answer is:

```
x = 0
sum = 0
while x < 1000:
    if (x%3==0 or x%5==0):
        sum += x
    x+=1
print(sum)
```

Under the progress page, the three awards that I most aspire to obtain are the following:

1. "C is for Commitment" – "Solve the first one hundred problems" – I wish to start slow by solving the first hundred problems before moving on to more difficult questions.

2. “As Easy As Pi” – “Solve questions 3, 14, 15 ...” – Pursuing this award seems like a fun way of motivating me to solve these questions.
3. “Gold Medal” – “the first to solve a problem” – I wish to learn how to be fast and efficient in thinking and coding.

Part (C)

The HIT that I have chosen is titled “Type the Text from the Images”. The worker is paid \$0.01 and no further details are provided. There are no qualifications, eligibility requirements or restrictions to do this task. The allotted time for this task is 10 minutes, and based on this, the hourly rate is \$0.06/hour. I think I can probably finish this task in around 2-3 minutes. In this case, the implied hourly rate is \$0.20/hour. This job expires in 7 days (25th November 2018). If 1 million people participated in this task, the HIT creator would spend \$ \$10,000.

Part (D)

I registered for kaggle and my username is josephinetancy. The title of the competition that I have chosen is “Quick, Draw! Recognition Challenge”. The task is to build a recognizer that can recognize and sort the millions of drawings that participants drew for the “Quick, Draw!” challenge. The sponsor of the competition is Google AI, an organization that conducts research in AI. Specifically, they aim to apply AI to new fields and develop tools to increase the accessibility of AI to the general public.

The submissions are evaluated according to the Mean Average Precision @ 3 (MAP@3):

$$MAP@3 = \frac{1}{U} \sum_{u=1}^U \sum_{k=1}^{\min(n,3)} P(k)$$

where U is the number of scored drawings in the test data, P(k) is the precision at cutoff k, and n is the number predictions per drawing. The 1st prize is \$12,000, 2nd prize is \$8,000 and the 3rd prize is \$5,000. There are two honor codes that each team/participant must abide to: firstly, each participant can only sign up for one Kaggle account and can only submit one file, and secondly there cannot be any sharing of code or data outside of teams. The entry and team merger deadline is November 27, 2018, 11:59 PM UTC (i.e. one must accept the competition rules to compete and join or merge teams before this date). The final submission deadline is December 4, 2018, 11:59 PM UTC. Each team can submit a maximum of 5 entries per day and can select up to 2 final submissions for judging. The submitted csv file should consist of the predictions of the test data, with at most 3 predicted word values (like “dog”, “crocodile”, “frog” etc.) for each key_id, according to the following format:

```
key_id,word
9000003627287624,The_Eiffel_Tower airplane donut
9000010688666847,The_Eiffel_Tower airplane donut
etc.
```

As Google aims to organize and increase usefulness and accessibility of information, Google will probably use the results of the competition to firstly, advance other pattern recognition AI models, specifically models that are used to recognize drawings. Secondly, the results of the competition can also be applied to other Google products like the Google search engine for

example, by including the additional function of searching for various handwritten line drawings in the search engine. More broadly speaking, holding this competition is also in line with the specific organization vision of Google AI written above, as conducting this competition also increase and develops the knowledge and accessibility of AI among the general public.

References:

Bell, Robert M., Yehuda Koren, and Chris Volinsky, All Together Now: A Perspective on the Netflix Prize," *Chance*, 2010, 23 (1), 24-29.