**Recommender System Algorithm**

**Melvin Ismanto**

**9018720198**

**Algorithm Summary**

My program uses the concept of collaborative-learning algorithm to predict the rating of movies in the test data using given MovieLens training data. I will be using one matrix factorization based algorithm, which is the Singular Value Decomposition (SVD), to do the prediction process. SVD is a factorization technique for matrix to generalize it into 3 matrix **U****UT**, with **U** and **UT**being the left and right singular (orthogonal) matrix respectively, and ****being a diagonal matrix. Note that I will be using the Surprise python library for this, to make use of its built-in SVD and accuracy functions. The main algorithm is as follows:

1. Scan the dataset from the input file.
2. Build the trainset and the testset separately.
3. Perform the SVD routine (fitting) to the user-item rating matrix. On a high-level description:
   1. Let vector of item and vector of user represented by `qi` and `pu` respectively, such that the dot product of those 2 vectors is the expected rating `rui`.
   2. We want to minimize the square error of this product, such that ⅀(rui-qiT. . pu)2 is minimum. We also added a regularization factor λ multiplied by the square sum of the magnitudes of user and item vectors so that it does not over-fit the training set, that is we add λ(||qi||2 + ||pu||2) to each squared error.
   3. To reduce the error between the predicted and actual value, the algorithm make use of some characteristics of the dataset, that is, for each user-item (u,i) pair we can extract 3 parameters. µ which is the average ratings of all items, `bi`which is the average rating of item i minus µ and `bu` which is the average rating given by user u minus µ, which makes the expected rating: rui = qiT . pu + µ + bi + bu
   4. Thus, the final equation to minimize is: minimum(p,q,bi,bu) ⅀(u,i)∈K (rui – qiT . pu - µ - bi – bu)2 + λ(||qi||2 + ||pu||2 bi2 + bu2)
   5. Using Stochastic Gradient Descent (SGD), we minimize the error of the equation with learning rate γ and number of epoch *n*.
4. Using the minimized singular matrix, predict the rating for the testset data.

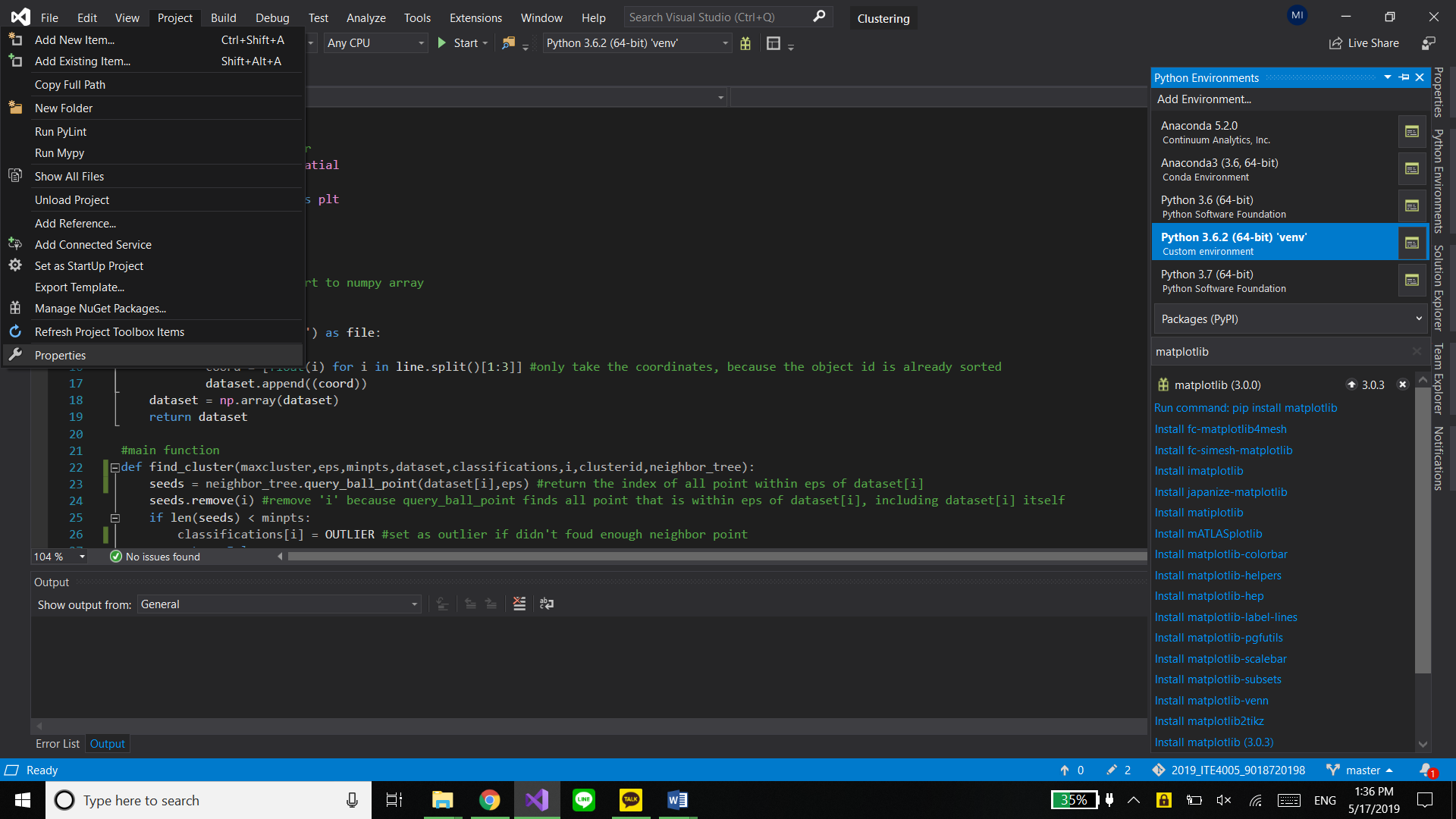
/////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////

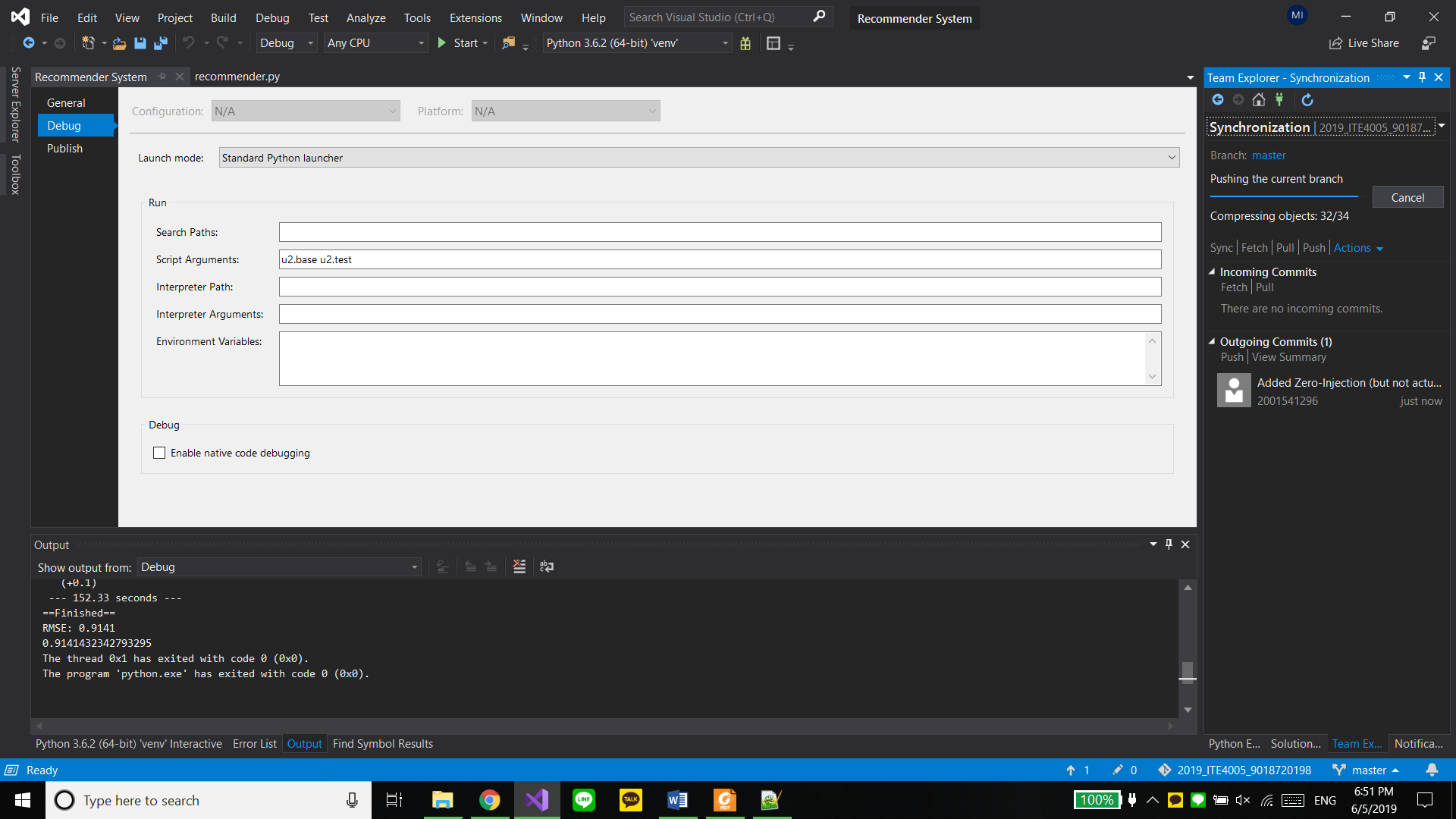
I uses the Microsoft Visual Studio Community **2019** (Version 16.0.1) IDE to write my code, using the Python version 3.6.2 language. So you can also use that to make compiling easier. When running the file, you should run the .sln file (project solution), not the .py file itself.

Download it here:

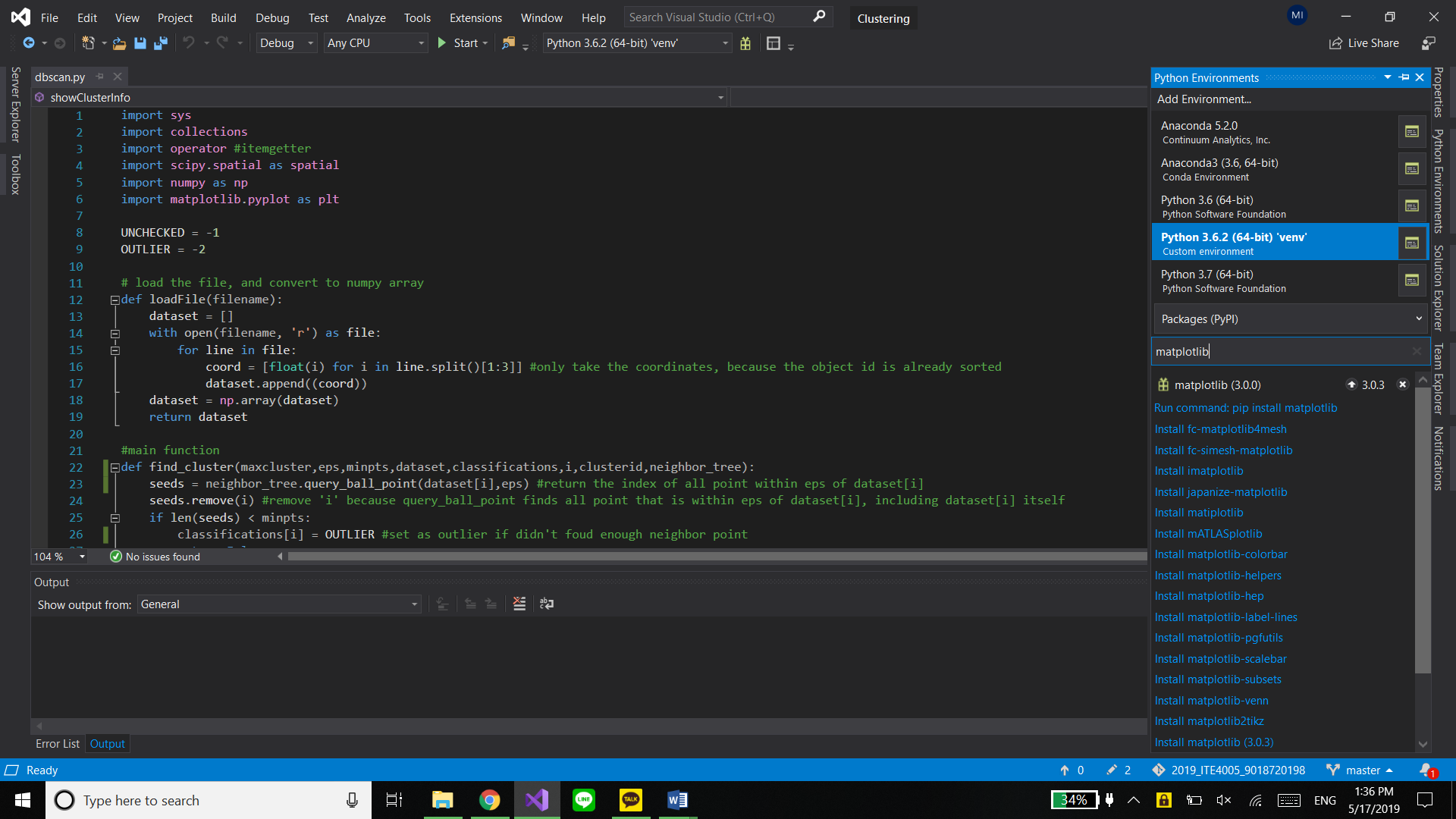
<https://visualstudio.microsoft.com/thank-you-downloading-visual-studio/?sku=Community&rel=16>

To change the argument parameter inside Visual Studio itself:





To add any package to the python environment, inside Visual Studio itself:



/////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////

This is how **I** install the Surprise python package

using pip:

type "pip install numpy"

then "pip install scikit-surprise"

These steps is to fix the "Microsoft Visual C++ 14.0 is required" error (which happens on my computer):

* go to : https://tinyurl.com/y4ho2g64 to download the visual studio installer
* run it and install the visual studio 2019 build tool
* go to the "installed" tab, and click on "modify" for "visual studio build tool 2019"
* on the right-hand side for "C++ build tools" install these things:
  + -MSVC v140 - VS 2015 C++ build tools (v14.00)
  + -Windows 10 SDK (10.0.16299.0)
* rerun the pip install command

These steps is to fix the "cannot run rc.exe" error (which also happens to me):

* go to: "C:\Program Files (x86)\Windows Kits\10\bin\10.0.16299.0\x86"
* copy the "rc.exe" and "rcdll.dll"
* paste to "C:\Program Files (x86)\Microsoft Visual Studio 14.0\VC\bin"
* rerun the pip install command

/////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////

Then just press F5 to start/debug the python program, or use it from the console command

I code this on a Windows 10 Pro 1809 (64 bit), with Intel i7-7500 CPU, and 16GB of RAM on my Laptop. It takes about 150 second on average for my program to analyze the provided text input file, and output the result to a text file.