**Division SIGMA - Resource Bank**

All resources below are for **supervised learning of text data** and are collected from **external sources** (mainly Youtube). The description under the links will inform you of how to use the resources properly for this competition, so please read them BEFORE accessing the resources. We also included videos about the concepts behind neural networks for those who are interested.

**Friendly Reminder:** If you think that the video is too lengthy, watch it at a **higher speed**.

| **Topic** | **Resource, Learning Duration, and Guides** |
| --- | --- |
| Python Download & Tutorial | **Python Setup Tutorial**  <https://www.youtube.com/watch?v=YYXdXT2l-Gg>  **➝ Learning Duration: 15 minutes**  This tutorial is detailed in the steps to download Python on both Mac and Windows. It also mentions several IDEs (Integrated development environment) - platforms in which you can write and run code. However, we encourage participants in this competition to use Google Colab instead of local IDEs, because it allows for collaborative coding. |
| **Python Download Link**  <https://www.python.org/downloads/>  **➝ Learning Duration: N/A**  This is the official Python download page. Pay careful attention to your operating system configurations (Mac, Linux, Windows - 32 bits or 64) |
| **Python Beginner Tutorial**  \* Reference to the released playlist |
| Python Refresher  (For people who have coded in Python before) | **Python 3 Basics Tutorial Series**  <https://www.youtube.com/playlist?list=PLQVvvaa0QuDe8XSftW-RAxdo6OmaeL85M>  **➝ Learning Duration: depends on the individual**  This tutorial series by YouTuber “Sentdex” offers many **short and concise videos** about basic concepts in Python programming. Please use this as a **refresher** if you have coded in Python before, which means that you should only watch the concepts that you are unfamiliar with to save time.  \* **All videos starting from #35 are mostly irrelevant to this competition**, you do not need to watch them. |
| Google Colab Tutorial | **2 ways to upload a CSV file to a Google Colab Notebook**  <https://medium.com/@simonprdhm/2-ways-to-upload-csv-files-to-google-colab-4d29ffa9db85>  **➝ Learning Duration: 10-15 minutes**  On the date of the competition, we will send the training data to each participant in the CSV (comma-separated values file) format. So it is useful to learn how to import a CSV using Pandas on Google Colab. |
| **Getting started with Google Colab | How to use Google Colab**  <https://www.youtube.com/watch?v=i-HnvsehuSw>  **➝ Learning Duration: 7 minutes**  Unlike the IDEs that people use on their local disks, Google Colab allows people to **collaborate when coding**. You can share your program just like sharing a Google Doc: simple and efficient! If you are unable to use Google Colab, don’t worry. You can still code your program on your IDE, but make sure to frequently update your code with your teammates. |
| GitHub Tutorial | **Getting started with GitHub | How to use GitHub**  <https://www.youtube.com/playlist?list=PLRqwX-V7Uu6ZF9C0YMKuns9sLDzK6zoiV> |
| Into to the concept of Machine Learning | **Brief Intro to the Concepts of Machine Learning**  <https://www.youtube.com/watch?v=f_uwKZIAeM0>  **➝ Learning Duration: 7 minutes** |
| Text processing options & Text data cleaning | **How to Clean Text Based Data for NLP - Part 3 - Python Yelp Sentiment Analysis**  <https://www.youtube.com/watch?v=5ndLJp_pb2Q>  **➝ Learning Duration: 12 minutes**  This video offers many text data cleaning techniques that are significant to Division Sigma’s competition (**hint!**). Some of the important techniques include stopwords, punctuations, lemmatization/stemming, tokenization, vectorization, Tf-Idf transform, etc. |
| Customization of models | \* For a **comprehensive and sequential course** about Machine Learning Models, please consider **Udacity's Intro to Machine Learning course** listed in the Extra Resource section on the last pages of the documents. |
| **Scikit-Learn Course - Machine Learning in Python Tutorial**  <https://www.youtube.com/watch?v=pqNCD_5r0IU>  **➝ Learning Duration: 1.5 - 3 hours (depends on your interest!)**  For Division Sigma’s competition problem, we encourage people to use the **Scikit-Learn** (as known as SKlearn) library with Python. This video by the YouTube channel “freeCodeCamp.org” covers some prominent **SKlearn AI models** that contestants should know. It has various timestamps to navigate through, in case the contestant doesn't want to watch the entire video. Contestants are **encouraged to finish Chapter 2**. Chapter 3 talks in-depth about some concepts behind AI, watch only if interested. \* The chapter divisions are shown in the actual video description. |
| **Scikit Learn Machine Learning SVM Tutorial with Python p. 2 - Example**  <https://www.youtube.com/watch?v=KTeVOb8gaD4>  **➝ Learning Duration: 30 minutes**  This video by YouTuber “Sentdex” focuses on the **Support Vector Machine** (SVM), which is one of the most classical Machine Learning models. The video exemplifies SVM in real-life situations and talks about the meaning of SVM parameters. |
| Pandas Tutorial | **Python Pandas Tutorial**  <https://www.youtube.com/playlist?list=PL-osiE80TeTsWmV9i9c58mdDCSskIFdDS>  **➝ Learning Duration: 6 - 7 hours**  This tutorial series with 11 videos by YouTuber “Corey Schafer” explains the basics of **Pandas**, a Python library dedicated to data manipulation. The series includes importing, reading, and writing data and simple row and column manipulation techniques which include grouping, moving and deleting, etc. |
| Conceptual Neutral Network | **Neural networks**  <https://www.youtube.com/playlist?list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi>  **➝ Learning Duration: 1 hour**  This playlist from the YouTube channel “3Blue1Brown” explains the conceptual knowledge behind neutral networks. |

Extra Resources

Most content in the extra resource is covered more concisely in the standard resource bank above. However, if you yearn for more knowledge and practice, relating to python or AI, check out the extra resources below.

| **Topic** | **Resource, Learning Duration, and Guides** |
| --- | --- |
| Intro to Python | **Python Complete Tutorials - All in one**  <https://www.youtube.com/watch?v=rfscVS0vtbw&t=39s> |
| **Python Complete Tutorials - Playlist**  <https://www.youtube.com/watch?v=oVp1vrfL_w4&list=PLQVvvaa0QuDe8XSftW-RAxdo6OmaeL85M> |
| Intro to Machine Learning, Text Preprocessing | **Udacity - Intro to Machine Learning**  <https://classroom.udacity.com/courses/ud120>  **\* FREE** course offered by Udacity, you would **NOT** need to pay!  This course introduces both **the concept and the code** for conventional machine learning algorithms. It has excellent explanations and practices about basic supervised classification models: Naive Bayes, Support Vector Machine (SVM), Decision Trees. It also explains how to preprocess input text before feeding it into an AI algorithm. Mini-projects in this course provide additional practice.  **➝ Learning Duration: 8 - 10 hours** (without mini-projects)  Since our competition is about **Supervised Classification of Text**, it is sufficient to just watch:   * Lesson 1 - Welcome to Machine Learning * Lesson 2 - Vectorization * Lesson 3 - SVM * Lesson 4 - Decision Trees * Lesson 11 - Text Learning * Lesson 15 - Evaluation Metrics   The mini projects in the course are great for practice. However, ignore them if you want to save time. For mini-projects, use Python 2 (preferably Python 2.7). Python 3 does NOT work! |