**Joseph Urie**

*Comp-Sci Undergraduate*

**CIS 467 - Term Project Final Report**

Poker with Python

**Abstract**

This project further investigates the topic of machine learning by demonstrating an ability to classify different data samples of playing cards. The machine developed during research can determine the poker value of a five card hand. The machine was trained and tested using provided data sets. The algorithm used during learning is support vector classification which bases heavily off of the support vector machine algorithm. The final machine is coded fully in Python (3.6.3) and incorporates the numpy, pandas, and Sci-kit learn libraries. Testing the machine on 1,000,000 different entries resulted in a 58% hit rate.

**Introduction**

Throughout the course, many artificial intelligence topics were introduced, but one of the most important was machine learning. The ability for a computer to learn on its own is a process even professionals in the field of computer science haven’t even mastered. What this project has accomplished has only dipped a toe into the undiscovered ocean of machine learning. Machine learning is a very important topic within the ever-growing field of artificial intelligence, even the great Bill Gates agrees. This motivated my exploration and decision for the project. Another key motivation was ECS 102 course taught at Syracuse University, my final project in that class was a simple poker game. I wanted to take that code and turn it into something much more complex with this project. While my initial goals set by my proposal haven’t been completely achieved, new goals have been set and much more is to be discovered.

**The Environment**

This project was performed under the instruction of Andrew Lee in the course CIS 467 Intro to Artificial Intelligence at Syracuse University. The source code for this project was created within Spyder, a python data science IDE. The code was written in Python version 3.6.3 and incorporates pre-build libraries. Numpy is used for its ability to handle arrays. Pandas is used to read in the data and .csv files. The Sci-kit Learn library is the largest library used and includes the algorithms and methods to train and test the machine with the data handled by pandas. Overall, Python was by far the easiest and most obvious choice as the language for this program. The Sci-kit Learn library includes anything I could ever need for machine learning on an investigation level and was built in Python. The ability to quickly master the language allowed for me to work harder on core parts of the machine rather than waste my time learning language syntax.

**The Goal**

The core work of this project is that the machine can read in 5 poker cards (an array of 10 numbers) and determine the poker value of the hand (a single number). The original data set was all 11 numbers. After splitting the data set into two seperate sets, we have a data set and a target set. Using the support vector classification algorithm, we fit it to the data. While training the machine, we only use the training data set. This data set is smaller with only about 25,010 entries. We give the machine both the data and the target set, so it can start looking for patterns and learning. Once the machine has information under its belt, we apply it to our test data set. The test data set is very large, standing tall at almost 1,000,000 entries. This larger test set has every possible combination of cards and can thoroughly examine the machine to see how well it works. That is exactly what we did and so far the machine can correctly guess 578,222 of those entries correct.

**Conclusion**

Machine learning is all about the learning. The machine has not completed learned the data and therefore isn’t 100% correct. Due to the growth of the field and my personal interests, I plan on further investigation with the code created for this project after the end of the course. The machine isn’t complete until it can correctly guess any hand of cards. I really enjoyed working on this project and designing this machine. There are definitely some things that would need to change if this project were to continue. My computer power wasn’t high enough and code took at least ten minutes to run even while working on minimum sized data. The rest of this project would have to take place in some sort of research environment. While there is more work to be done, this project managed to open a lot of doors into machine learning’s future.

**Appendix**

*Project Proposal*

Poker with Python

**Problem Description**

This project will use machine learning and python, more specifically sklearn. The final product will use regression, classification, and clustering to examine a poker hand data set originally created by Franz Oppacher and Robert Cattral. Once the machine gets comfortable with the data set, it, in theory, should be able to outplay a human player in a series of poker/card games including Texas hold’em and five card draw.

**Motivations**

Machine learning is arguably one of the most important parts of artificial intelligence. I personally created a very simple poker game in java code my freshman year and noticed a lot of problems with its efficiency being that the machine was just too predictable. A simple card game will be the quickest way for a machine to learn and make it easy for humans to interact with this new machine.

**Nature of the Proposed Work**

This project would be considered an applied project. Personally, applied projects tend to be the most rewarding to finish as well as the ones I learn the most from. Applying what I know is a great way to test whether or not I actually know it. While the project relies heavily on theories learned in class, the final product of this project will be working code.

**Methods**

* Python – The project in its entirety will be written in the python coding language. There are lots of machine learning libraries readily available for users at absolutely no cost. It is easy to practice machine learning in this sort of environment. Python is available for everyone and can be downloaded from their website: <https://www.python.org/downloads/>
* Scikit-Learn – Also known as sklearn, it is a machine learning library for python. It is open source and accessible to anyone willing to use it. It includes simple and efficient tools for data mining and data analysis. It is built on prior python libraries including NumPy, SciPy, and matplotlib. It can be accessed from its homepage: <http://scikit-learn.org/stable/>
* Poker Hand Data Set – This free to download dataset is provided by the UCI Machine Learning Repository. We will be using this data set to try and teach our machine to become familiar with cards and the game of poker. Each record is an example of a hand consisting of five playing cards drawn from a standard 52 card deck. Each card has two attributes, suit and rank. The data set also recognizes poker hands and their winning values. This specific data set can be downloaded from: <http://archive.ics.uci.edu/ml/datasets/Poker+Hand>

**Timeline**

|  |  |  |
| --- | --- | --- |
| **Date:** | **Milestone:** | **Demonstration:** |
| 10/27 | Initial project ideas have been researched and a final topic has been decided. Term Project Proposal Submitted. | Term Project Proposal |
| 10/31 | Personal Computer has been set up with a fresh, up to date install of Python, a stable release of sklearn, and the poker data set. | n/a |
| 11/17 | The first compliable prototype of the code will be finished. Debugging will begin after this. | Prototype #1 – Source Code  *Project demo will be available* |
| 11/27 | Second prototype should be completed. This prototype should have little to no errors and look very similar to what will be the final product. | Prototype #2 – Source Code  *Project demo will be available* |
| 12/6 | Final Product will be completed and be presented to the class. | Final Product – Source Code  *Demo will be presented to entire class during lecture.* |