**Student Name- Ayush Ghai**

**Student Id-18302819**

**Email Address- ghaia@tcd.ie**

**Proposed Supervisor- Mr. Glenn Strong**

1. ***Introduction***

Machines have become a part and parcel of our life. Computers, mobile phones are some machines which have now become a basic need of human. What goes behind the creation of these machines is the thousands of lines of code in numerous different languages. Earlier, learning to code was not a very common thing as it is today. There were very limited resources from where people can learn to code. Nowadays there are a lot of platforms where one can learn to code like HackerRank, HackerEarth and Scratch.

Scratch is a platform that is different from others as it is primarily targeted at children where they can learn the concepts of programming in an interactive and fun way which can then be applied to python and java programming languages. Scratch allows the users to put their work on their repositories and remix or use someone else’s code in their project. This platform helps the children to learn coding but also raises an issue of whether the children involved in these platforms learning or they are copying another user’s code. Also, another question that arises is whether there is any change in sophistication of user’s code as they code for a while.

I propose to find out more about the change in sophistication of programming of a scratch user over the years. For this I propose to analyze the data till 2012 and scrape some of the data after 2012 and look for patterns that can answer my questions.

1. ***Research Question***

Scratch- Is there a change in sophistication of code of scratch users over the years?

1. ***Research Aims***

To find out whether there is a change in sophistication of code of scratch users as they grow in age.

1. ***Motivation***

I have been practicing coding on different platforms like HackerEarth and Hackerrank for last few years and it has helped me a lot in learning new languages and my coding has changed over the years by using these platforms. I read about scratch made by MIT which helps children of age 12 to 16 learn to code using blocks and then these programming concepts can be implemented in languages like python and java. This is a good initiative by MIT and has attracted a lot of users of different age groups to come and work on that platform. This platform is also used by the users to store their codes and they can also use the code of other users to remix and make a new and better code. My supervisor provided me with a few papers [1][2][3] to read in which they have scraped data from scratch website and analyzed that data to generate insights. This motivated me to think more about this platform and raised a question in mind. I propose to find out the change in sophistication of code of scratch user as they grow in age.

1. ***A list of Sources of literature***
2. Efthimia Aivaloglou , Felienne Hermans , Jesús Moreno-León , Gregorio Robles, A dataset of scratch programs: scraped, shaped and scored, Proceedings of the 14th International Conference on Mining Software Repositories, May 20-28, 2017, Buenos Aires, Argentina
3. Deborah A. Fields , Yasmin B. Kafai , Michael T. Giang, Youth Computational Participation in the Wild: Understanding Experience and Equity in Participating and Programming in the Online Scratch Community, ACM Transactions on Computing Education (TOCE), v.17 n.3, p.1-22, August 2017
4. Leonel Morales Diaz, Luis Felipe Ayala Lopez, A Classification of Programming Styles in Scratch, Proceedings of the 8th Latin American Conference on Human Computer Interaction Article No. 9 Antigua Guatemala, Guatemala
5. ***Proposed methodology to answer the research question***

First, not all the data is available to use and analyze for my research. So, the first task will be scraping that data from scratch. Then the second step that I propose to do is cleaning that data to handle the data that is not useful to me. Then use this data to cluster and divide the users on the basis of the code that they have written and analyzing whether there is any change in sophistication of their code as they grow in age.

1. ***Description of your area of research which places your research in context***

Using data science for generating insights that can be used in educational domain is what my research question is targeted.

1. ***Evaluation criteria and method***

There is a change in code of the user as he grows as a coder. As we start coding, our codes are very basic without any optimisation we usually use longer codes and unoptimized codes and we grow as a coder we tend to optimise the code and learn about new functions and techniques of coding. This change in coding of the users will help us to analyse whether that user is learning on the platform or not. The change in the way a user codes will help us to generate a pattern and hence answer my question.

1. ***Ethics***
2. ***Potential Benefits of your Research***

The research I am proposing will help us to analyse the benefits of these online coding platforms. It will also help us to analyse what changes need to be done in these coding platforms to make it more suitable for the users. It can help us analyse the way in which the users of age 12-16 are thinking.

1. ***Your skills audit including relevant skills you have mastered and skills you would like to develop as part of this piece of research. For the skills you would like to develop***

I have done something related to knowledge that is to be implemented in the research during my bachelor studies that involved mining data and doing its opinion mining that will be helpful for me in this research. I am new to data analysis and I would love to gain data analysis skills and using codes and tools to analyse the data and generating insights from that data.