**Course: Computer Science Module: 6006CEM Machine Learning and Related Applications**

**Student Name: Saad Iftikhar**

**Student ID: 9789180**

**Coventry GitHub Repository URL** or **Coventry OneDrive URL** (mandatory):

< <https://github.coventry.ac.uk/iftikhars/9789180-SI-s1>>

**Table of contents**

[Table of figures 2](#_Toc84718354)

[Introduction 3](#_Toc84718355)

[**Problem statement** : Predicting the Suicide rates of males and females at different ages 3](#_Toc84718356)

[Motivation: 3](#_Toc84718357)

[Related work 3](#_Toc84718358)

[Existing approaches or methods and their results 3](#_Toc84718359)

[Similarities and differences between this work and the existing works 3](#_Toc84718360)

[Dataset 5](#_Toc84718361)

[Pre-processing 6](#_Toc84718362)

[Implementation 7](#_Toc84718363)

[Results 8](#_Toc84718364)

[Evaluation 9](#_Toc84718365)

[Conclusion 10](#_Toc84718366)

[References 11](#_Toc84718367)

[Appendix A 11](#_Toc84718368)

[Appendix B 13](#_Toc84718369)

# Table of figures

# Introduction

## **Problem statement** : Predicting the Suicide rates of males and females at different ages

Suicide is a significant cause of mortality globally, resulting in the loss of a substantial number of  people lives. There is currently no agreement on the optimum performance qualities needed to use suicide risk prediction algorithms in clinical practise [(Sanderson et al., 2020)](#Sanderson2020). As suggested by [Lutz, Mechawar and Turecki (2017)](#Lutz2017) studies in recent times, in neuroanatomy, genetics and molecular psychiatry have produced models in which behavioural dysregulation, including suicidal behaviour, develops as a function of biological adaptation in major brain systems. Being a male is without a question most common risk factor for suicide across the world. Male suicide is more over twice as prevalent as female suicide globally [(Large, 2018)](#Large2018). Biological aging could contribute to suicide risk as suggested by the relationship between aging and rising suicide rates [(Conwell Raby, and Caine 1995)](#Conwell1995). Our ability to predict suicide has remained at a chance level for decades. Potentially accurate measurable and effective detection of suicide risk may be possible because of machine learning [(Linthicum, Schafer, and Ribeiro 2019)](#Linthicum2019).

## Motivation:

As suggested by the above research that suicide rates increase with the increase in age and the likelihood of suicide is substantially more prevalent in human males as opposed to females. The motivation behind this report is to study at which age females and males are most likely to commit suicide. This report focuses primarily on aging and sex using machine learning classification model to predict at which age a person with a specific gender is most likely to commit suicide.  
  
The desired outcome of this report is to present results that could be used in further research to determine that if there are any biological factors that increase the risk of suicide at a certain age by gender.

## Related work

## Existing approaches or methods and their results

There was a similar study that was carried out by [(Ritchie et al., 2015)](#Richie2015) in which they also used data set from the World Health organization to predict suicide rates per 100,000 people overall and by age group and sex. A part of their research included suicide rates by age and suicide rates by gender. Their research showed that older people were more likely to commit suicide then younger people. Their research also concluded that men are more likely to commit suicide then women.

There are 94 different works available on the same Kaggle [dataset (Szamil, 2018)](#szamil2018). The users AAKAR MUTHA [(MUTHA, 2020)](#Muta2020) Adeline [(Adeline, 2020)](#Adeline2020) and others have conducted research using the mentioned dataset [(Szamil, 2018)](#szamil2018) but the focused data in their research was different to the focal data of this research. The most similar research to this project was done by the Kaggle users [(Eralp, 2018)](#Eralp2018) and [(PHILAVONG, 2018)](#Philavong2018).

## Similarities and differences between this work and the existing works

[(Eralp, 2018)](#Eralp2018) perform their research using a regression model. Their work was only done on data from one country, and they included an age column to determine if the rate of suicide was increasing or decreasing with respect to time. [(PHILAVONG, 2018)](#Philavong2018) checks different age ranges to check that if the mentioned age range can affect a person into committing suicide. They check the validity of their model using statistical techniques other than regression models.

[(Ritchie et al., 2015)](#Richie2015) has a far more comprehensive study them the two mentioned earlier but a part of their analysis does focus on age, gender, suicide, and their interrelationship but this study also focuses on that is the rate of suicide was increasing or decreasing with respect to time.

This project will only focus on two related things which are likelihood of suicide by age and gender and the rate of suicide by age and gender in each corresponding population. The similarities this project has with the mentioned projects is the affect of age and gender on suicide. This project is most like the project by [(PHILAVONG, 2018)](#Philavong2018) but it will use data filtering and pre-processing techniques which were not used by [(PHILAVONG, 2018)](#Philavong2018). This project is different from all mentioned projects as it uses classification Models which are not used in any of the mentioned projects. In that sense it is most similar to [(Adeline, 2020)](#Adeline2020) work which also uses classical modelling using the same dataset [(Szamil, 2018)](#szamil2018) but their analysis is entirely on suicide in relation to population.

# Dataset

The dataset used in this project has been taken from Kaggle from the user Szamil [(Szamil, 2018)](#szamil2018). The data belongs to World Health Organization (WHO).

# Pre-processing

# Implementation

# Results

# Evaluation

# Conclusion

# References

Large M. M. (2018). The role of prediction in suicide prevention. *Dialogues in clinical neuroscience*, *20*(3), 197–205. <https://doi.org/10.31887/DCNS.2018.20.3/mlarge>.

Conwell, Y., Raby, W. N., and Caine, E. D. (1995). Suicide and aging. II: The psychobiological interface. *International psychogeriatrics*, *7*(2), 165–181. <https://doi.org/10.1017/s1041610295001955>.

Lutz, P. E., Mechawar, N., and Turecki, G. (2017). Neuropathology of suicide: recent findings and future directions. *Molecular psychiatry*, *22*(10), 1395–1412. <https://doi.org/10.1038/mp.2017.141>.

Linthicum, K. P., Schafer, K. M., and Ribeiro, J. D. (2019). Machine learning in suicide science: Applications and ethics. *Behavioral sciences & the law*, *37*(3), 214–222. <https://doi.org/10.1002/bsl.2392>.

Mutha, A. (2020, July 23). *Comparison of differnet Regression Models*. Kaggle. Retrieved October 9, 2021, from https://www.kaggle.com/aakarmutha/comparison-of-differnet-regression-models.

Adeline, N. (2020, May 15). *Natasha\_visualizations*. Kaggle. Retrieved October 9, 2021, from https://www.kaggle.com/natashaadeline/natasha-visualizations.

Szamil. (2018, August 29). *Who suicide statistics*. Kaggle. Retrieved October 5, 2021, from https://www.kaggle.com/szamil/who-suicide-statistics.

Eralp, C. (2018, December 2). *Practice #2: Suicide analysis with regression*. Kaggle. Retrieved October 9, 2021, from https://www.kaggle.com/cengizeralp/practice-2-suicide-analysis-with-regression.

PHILAVONG, N. O. N. (2018, December 6). *Age and suicide*. Kaggle. Retrieved October 9, 2021, from https://www.kaggle.com/monethong/age-and-suicide/notebook.

Sanderson, M., Bulloch, A. G. M., Wang, J. L., Williams, K. G., Williamson, W., & Patten, S. B. (2020, February 18). Predicting death by suicide following an emergency department visit for parasuicide with administrative health care system data and machine learning. thelancet. Retrieved October 9, 2021, from https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(20)30025-0/fulltext.

Ritchie, H., Roser, M., & Ortiz-Ospina, E. (2015, June 15). Suicide. Our World in Data. Retrieved October 9, 2021, from https://ourworldindata.org/suicide.

**Later cross argument if required: https://wonder.cdc.gov/controller/datarequest/D76**

# Appendix A

< A suggested checklist for you, for full details please refer to the coursework brief >

1. The following naming convention is used for the Coventry GitHub Repository and Coventry OneDrive

StudentID-Initials-s1

For example, for a student Alan Turing whose student ID was 1234567, it should be

1234567-AT-s1

Failing to follow the naming convention may delay the release of marks and feedback for your coursework.

1. **Coventry** GitHub Repository URL **or** **Coventry** OneDrive URL: added to the top of this report
   1. Coventry GitHub Repository includes:

* URL of the selected dataset(s) included in README
* The original selected dataset(s)
* Source-code (.ipynb)
* Demonstration video (.mp4)
  1. Coventry OneDrive folder includes:
* URL of the selected dataset(s) included in a separated text file
* The original selected dataset(s)
* Source-code (.ipynb)
* Demonstration video (.mp4)

1. Source-code added **as text** in Appendix B (below)
2. Submission in the form of a **Word** document. *\*\*Other format is not accepted.*

# Appendix B

< **Replace** this instruction with all the Programming Code for the coursework.

Make sure you have highlighted and referenced any code not written by you >

< **DO NOT** use screenshots of your code here. Your code should be presented **as text**.

There are many good tools to help you format your code such as <http://hilite.me> >

< You can select and copy **all code at once** in a notebook by:

1. Graphical user interface, text, application

   Description automatically generatedclicking in any cell of the notebook, the cell will be highlighted in green as below
2. Graphical user interface, text, application

   Description automatically generatedthen press Esc on your keyboard, the selected cell will be highlighted in blue as below
3. now you can Ctrl+A to select all cells of the notebook

Graphical user interface, text, application

Description automatically generated

1. and then copy and paste as normal to some tool, such as hilite.me above, make sure you select a correct language (Python), then click Hightlight

Graphical user interface, text, application

Description automatically generated

1. you now can select the text in the Preview and copy and paste it over to this Appendix

Graphical user interface, text, application, email

Description automatically generated

1. finally, remember to remove all text in this instruction for this Appendix >