**CSCI 5931 Research topics in Computer Science**

Project report of Predicting Happiness using key factors

From Mental Health DataSet

Instructor: Prof. Wu, Yalong

Student Name 1: Monish Kamal Bathini

Student Name 2: Ajay Kumar Reddy Poreddy

Student Name 3: Deekshith Machidi

Student Name 4: Giridharan Ilango

**1.Introduction**

* 1. **Project Introduction**

According to the World Health Organization released report in 2022 (World Mental Health Report: Transforming Mental health for All), around 970 million people around the world (~13% of the world's population) were living with mental disorders. This surprising statistic underscores the urgency of addressing mental health challenges on a global scale. The impact of mental health disorders extends beyond individual suffering; it also has serious economic consequences. It was reported that mental health disorders had an economic impact worth approximately over 800 billion USD in 2010, and this number is estimated to double by 2030, placing a significant strain on economies worldwide.

Depression stands out as one of the largest and most dangerous among all mental disorders. A surprising 29% of the global population is affected by this tiring condition. What makes depression even more concerning is its ability to adversely affect chronic health conditions, such as cardiovascular disease, cancer, diabetes, and obesity. It’s complicated aetiology and chronic clinical features make it difficult for individuals to be consciously aware of their own depressive emotions, often leading to severe consequences, including suicide.

Looking ahead, it is projected that in the coming 20 years, depression will become the leading cause of disability in high-income nations. The World Health Organization (WHO) now ranks major depression as one of the most burdensome diseases in the world, further emphasizing the critical importance of addressing this issue. To reduce the economic, physical health, and mental health losses associated with depression, it is imperative to develop innovative solutions for early detection and treatment.

Traditionally, depression diagnosis has relied on a psychiatrist's subjective analysis of questionnaires provided to patients. However, depression is characterized by a lack of clear boundaries or specific symptoms, making human intervention essential for accurate diagnosis. The alarming reality in low-and-medium income countries is that there is typically only one psychiatrist available to serve 200,000 or more people. This surprising ratio highlights the limited resources available to tackle mental health disorders in regions where they are most needed. Additionally, mental health services in these areas are often inaccessible or financially out of reach for many individuals.

In response to these challenges, this research project is assured to provide a groundbreaking solution for diagnosing depression without the need for human intervention. By using the potential of big data analytics, we aim to create an automated depression detection system that can revolutionize mental health care worldwide. This innovative approach has the potential to improve early diagnosis, increase access to care, and ultimately alleviate the economic, physical, and mental health burdens associated with depression. In the pages that follow, we will delve deeper into the methodology and potential benefits of this pioneering research attempt.

* 1. **Using this Project**

The major purpose of this project is to address the global mental health by building an innovative model using big data analytics, aiming to improve early diagnosis, increase access to mental health care, and reduce the significant economic, physical, and mental health burdens associated with depression worldwide.

* 1. **Prerequisite Knowledge for Beginners**

If you want to understand what this report is talking about, you need to have a basic knowledge of global mental health and conditions like depression and also an introductory understanding of big data analytics and machine learning contributes to grasping the potential of the model developed to address these global mental health challenges.

* 1. **Report Structure**

In the next section of this report, we will provide an in-depth exploration of the design associated with depression prediction. The focus will be the systematic adding of the map-reduce into the model, explaining how this framework is methodically employed to extract specific insights crucial for accurate depression prediction. Simultaneously, we will delve into the use of visualization techniques within the model, aiming to explain the dataset's patterns and trends visually.

**2. Design and Implementation**

**2.1 Data**

The data we used is the data from “GSS Data Explorer” which includes both demographic information and respondents' [opinions](https://en.wikipedia.org/wiki/Opinion) on matters ranging from government spending to the state of [race](https://en.wikipedia.org/wiki/Race_(classification_of_human_beings)) relations to the existence and nature of [God](https://en.wikipedia.org/wiki/God). Because of the wide range of topics covered and the comprehensive gathering of demographic information, survey results allow social scientists to correlate demographic factors like age, race, [gender](https://en.wikipedia.org/wiki/Gender_survey_question), and urban/rural upbringing with beliefs and thereby determine whether, for example, an average middle-aged black male respondent would be more or less likely to move to a different U.S. state for economic reasons than a similarly situated white female respondent; or whether a highly educated person with a rural upbringing is more likely to believe in a transcendent God than a person with an urban upbringing and only a high school education.

**2.2Pre-Processing**

The dataset contains a total of 28 columns/features but in some columns, the unknown data is comparity high rather than the known inorder to compensate this we which columns are so and such columns are either removed or filled with relevant numerical/categorical values as shown below,

A screenshot of a computer

Description automatically generated

Counting the number of null values

A white background with red text

Description automatically generated

Removing unnecessary columns

A screenshot of a computer

Description automatically generated

Replaced the Missing values

**2.3 Visualization**

**Graphs:**

**A graph with different colored squares

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer screen

Description automatically generated**

**A graph of a divorce status

Description automatically generated with medium confidence**

**A pie chart with different colored circles

Description automatically generated**

**A graph of a number of people

Description automatically generated with medium confidence**

**2.4 Model Building**

**A screenshot of a computer program

Description automatically generated**

**A number of numbers and symbols

Description automatically generated with medium confidence**

**A screenshot of a computer program

Description automatically generated**

**A screenshot of a document

Description automatically generated**

**2.5 Using MadReduce**

**A screenshot of a computer program

Description automatically generated**

**A screenshot of a computer program

Description automatically generated**

**3. Conclusions**

**3.1 Summary**

In conclusion, this project represents a significant step towards addressing the global mental health, particularly focusing on the impact of depression. The need to develop an innovative depression prediction model, leveraging big data analytics and map-reduce, results in the betterment of the society. The urgency of the issue is highlighted by shocking statistics revealing the widespread impact of mental disorders and the projected economic burden of depression. The challenges associated with traditional methods, particularly in regions with limited mental health resources, further emphasize the need for accessible and efficient solutions.

**3.2 Future Work**

As we move forward, the future work for this project involves refining and optimizing the depression prediction model. Continuous validation and improvement will be pursued to enhance the accuracy and reliability of the model. Additionally, efforts will be directed towards the scalability of the model, ensuring its applicability across diverse populations and contexts. Continuous engagement with emerging technologies and advancements in big data analytics will be essential for staying updated in the field of mental health. Collaborative efforts with healthcare professionals, researchers, and policymakers will be crucial to ensure the effective implementation and global impact of the developed model. Ultimately, the ongoing evolution of this project aims to contribute significantly to early diagnosis, increased accessibility to mental health care, and the reducing of the bad effects of depression on a global scale.

**4. References:**

[1] J. A. Naslund *et al.*, "Digital Innovations for Global Mental Health: Opportunities for Data Science, Task Sharing, and Early Intervention," (in eng), *Curr Treat Options Psychiatry,* vol. 6, no. 4, pp. 337-351, Dec 2019, doi: 10.1007/s40501-019-00186-8.

[2] T. C. Russ *et al.*, "How data science can advance mental health research," (in eng), *Nat Hum Behav,* vol. 3, no. 1, pp. 24-32, Jan 2019, doi: 10.1038/s41562-018-0470-9.

[3] E. Saravia, C. H. Chang, R. J. D. Lorenzo, and Y. S. Chen, "MIDAS: Mental illness detection and analysis via social media," in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 18-21 Aug. 2016 2016, pp. 1418-1421, doi: 10.1109/ASONAM.2016.7752434.

[4] R. V. Saveanu and C. B. Nemeroff, "Etiology of depression: genetic and environmental factors," (in eng), *Psychiatr Clin North Am,* vol. 35, no. 1, pp. 51-71, Mar 2012, doi: 10.1016/j.psc.2011.12.001.

[5] C. Sudlow *et al.*, "UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age," (in eng), *PLoS Med,* vol. 12, no. 3, p. e1001779, Mar 2015, doi: 10.1371/journal.pmed.1001779.

[6] S. Yang, P. Zhou, K. Duan, M. S. Hossain, and M. F. Alhamid, "emHealth: Towards Emotion Health Through Depression Prediction and Intelligent Health Recommender System," *Mobile Networks and Applications,* vol. 23, no. 2, pp. 216-226, 2018/04/01 2018, doi: 10.1007/s11036-017-0929-3.

[7]. Muennig, Peter (October 6, 2011). ["The general social survey-national death index: an innovative new dataset for the social sciences"](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3199263). *BMC Research Notes*. **4** (385): 1–4. [doi](https://en.wikipedia.org/wiki/Doi_(identifier)):[10.1186/1756-0500-4-385](https://doi.org/10.1186%2F1756-0500-4-385). [PMC](https://en.wikipedia.org/wiki/PMC_(identifier)) [3199263](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3199263). [PMID](https://en.wikipedia.org/wiki/PMID_(identifier)) [21978529](https://pubmed.ncbi.nlm.nih.gov/21978529).