

Final project idea proposal.

Domestic violence towards women in Kazakhstan and the world.

Zhakanova Asem, IT-2201

Vitalii Khan, IT-2201

Zharkenova Aruzhan, IT-2201

Problem or idea description:

Domestic violence is a widespread issue that affects numerous individuals worldwide, crossing geographical, cultural, and socioeconomic boundaries. Addressing this problem systematically is crucial due to its prevalence. While there is available data on domestic violence, the challenge lies in synthesizing and understanding the patterns, contributing factors, and variations across different regions. This complexity requires a comprehensive approach to analyze and derive meaningful insights from the dataset.

Our project aims to utilize a dataset encompassing incidents of domestic violence worldwide to contribute to a deeper understanding of this critical issue. Through the use of advanced data analysis, machine learning techniques, and visualization tools, we aim to extract valuable insights that can inform policymakers, researchers, and organizations working towards mitigating the impact of domestic violence.

Background information on the problem or idea:

Violence against women has a global impact, affecting women of all ages, classes, races, and ethnicities. Recent estimates indicate that 30 percent of women aged 15 or older worldwide have experienced physical and/or sexual intimate partner violence during their lifetime. This type of violence is the leading cause of homicide death in women globally and has significant health consequences. The economic and social costs associated with violence against women are substantial, and global evidence demonstrates that it consistently hinders development efforts at various levels, leading to the depreciation of physical, human, and social capital. Our goal in working on this project is to contribute to the improvement of the situation of domestic violence around the world.

Available solutions with links:

1. We're extracting data from a government website by crafting custom web scraping tools for each site, navigating dynamic content with advanced algorithms, all while making sure we stay within legal and ethical boundaries for data extraction.

<https://3idatascrapingblog.medium.com/how-to-extract-data-from-any-government-website-b7027897b187>

2. Creating your own Telegram bot is achieved through meticulous development. We construct tailored solutions, utilize Telegram's API, and provide a systematic guide to ensure the creation of your bot is a methodical and accessible process.

<https://medium.com/spidernitt/how-to-create-your-own-telegram-bot-63d1097999b6>

3. Seamlessly integrating ChatGPT with Telegram requires a strategic approach. We develop customized solutions, leverage Telegram's API, and offer a systematic guide to ensure a smooth and efficient integration process, facilitating pleasant interactions between users and our bot.

<https://medium.com/@mirceaaiosif/how-to-integrate-chatgpt-with-telegram-8b8d6075f955>

4. Utilizing this guide empowers us to effectively implement BERT for everyone. We custom-tailor solutions, leverage the capabilities of BERT, and follow a systematic tutorial to ensure a smooth integration. This approach helps us to enhance our project.

<https://www.kaggle.com/code/harshjain123/bert-for-everyone-tutorial-implementation>

5. Optimizing Data Manipulation in Python with Pandas: A Practical Tutorial and Implementation Guide. We apply this guide to better our proficiency in data manipulation, utilizing different solutions, and following a systematic approach to ensure effective implementation. This empowers our project with good data processing capabilities using Pandas.

<https://www.geeksforgeeks.org/data-manipulation-in-python-using-pandas/>

6. By following this guide, we use the power of diverse Python libraries, implementing custom solutions and a systematic approach. This enriches our project, enabling sophisticated data manipulation through the utilization of these key libraries.

<https://clouddevs.com/python/libraries-for-data-manipulation/>

7. We efficiently delve into the fundamentals of machine learning, implementing custom solutions and a systematic approach. This enriches our project, allowing for a robust understanding and practical application of machine learning using TensorFlow."

<https://www.tensorflow.org/resources/learn-ml/basics-of-machine-learning?hl=ru>

How to get the data?

Brief description of your solution

Tech stack that will be used

1. Introduction:

For this project, our goal is to implement a scientific paper using Google Colab, integrating advanced techniques such as the BERT model alongside Python for data manipulation, establishing a database, and leveraging TensorFlow for machine learning. The initial dataset, collected from various sources, will be seamlessly integrated into the Google Colab environment, allowing for a structured approach to data handling. In addition to these elements, we are planning to enhance user interaction and accessibility by implementing a Telegram bot related to this topic. The bot will not only provide information but will also offer additional functionalities

2. Data Retrieval and Exploration:

To commence the project, we will employ Python's pandas library to retrieve data from diverse sources and load it into a dataframe within Google Colab. This initial step facilitates an exploration of the dataset, ensuring a clear understanding of its structure and content. Data cleaning or preprocessing will be conducted as necessary.

3. Database Integration:

Acknowledging the scale and persistence of the data, our plan involves establishing a database to centralize information. Google Colab will be connected to the database, creating a seamless bridge between the Colab environment and data storage.

4. Importing Data into Google Colab:

If needed, we will use pandas to import cleaned and preprocessed data directly into Google Colab. This ensures data availability and facilitates various analyses or machine learning tasks within the Colab environment.

5. Exploring Data Using External Websites:

For a comprehensive analysis, our approach includes exploring the dataset using external websites. Employing various tools, we aim to extract insights such as statistics and trends, enhancing the depth of our analysis.

6. Utilizing BERT Model for NLP Tasks:

As a significant component of our tech stack, we will incorporate the BERT model for natural language processing (NLP) tasks. We find this model quite useful for our project, and it will enhance our ability to derive nuanced information and context from textual data within the project.

7. Tech Stack:

Our tech stack encompasses Python for scripting, pandas for data manipulation, TensorFlow for general machine learning tasks, and the BERT model for NLP tasks. Google Colab serves as our primary environment for implementation.

8. Conclusion:

In summary, this project covers the entire lifecycle of implementing a scientific paper using Google Colab. Starting from data retrieval and exploration to database integration and utilizing external websites for in-depth analysis, the inclusion of the BERT model further enhances our capabilities in natural language processing, providing a robust foundation for a successful project outcome.