EPIDEMIC PREDICTION APP USING MACHINE LEARNING

Aparajita Goswami

18/11/2023

Problem Statement

Through this project, our goal is to create a predictive application using machine learning principles, that can identify the early signs and risk factors associated with potential epidemics. This predictive application would be invaluable for our community, allowing for proactive measures, resource allocation, and timely responses to mitigate the impact of infectious disease outbreaks. The successful development and deployment of this idea can significantly enhance public health preparedness and community well-being.

Market/Customer/Business Need Assessment

The idea addresses critical market, customer, and business needs related to public health, data analysis, risk mitigation, community well-being, and global health security. By developing a machine learning app for epidemic detection, the project has the potential to make a significant positive impact on various stakeholders while creating business opportunities in the healthcare and data analysis sectors.

1. **Public Health Preparedness:**
    - **Market Need:** There is a growing concern for public health, especially in the face of emerging infectious diseases. The recent experiences with pandemics like COVID-19 have highlighted the need for early detection and response to epidemics.
    - **Customer Need:** Health authorities, governments, and communities require advanced tools to monitor and predict epidemics, enabling them to take proactive measures to safeguard public health.
    - **Business Opportunity:** Developing an effective epidemic detection system addresses a critical market need, with potential customers ranging from healthcare institutions, government agencies, to international organizations.

2. **Risk Mitigation:**
    - **Market Need:** The ability to predict epidemics before they reach critical levels is crucial for reducing their impact on public health and minimizing economic disruptions.
    - **Customer Need:** Customers need tools that enable them to take proactive measures, such as resource allocation, vaccination campaigns, and travel restrictions, to mitigate epidemic risks.
    - **Business Opportunity:** Providing predictive models for epidemic detection can create opportunities for consultancy services, software development, and partnerships with governments and international organizations.

Target Specifications and Characterization

The proposed machine learning idea will benefit health activists, specialists, and the community as a whole, as the model will be able to detect an epidemic in its early stages, and hence the risk mitigation process can be started at the right time. This will be crucial for the economy as well, as resource allocation, and healthcare at the right time, will cut down several emergency and casualty prices.

External Search:

1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9101183/#:~:text=SEIR%20Model,as%20lockdowns)%20may%20be%20effective.
2. https://www.psu.edu/impact/story/forecasting-infectious-diseases/
3. https://www.psu.edu/impact/story/forecasting-infectious-diseases/
4. https://ieeexplore.ieee.org/document/10039594/

Data for Small Scale Implementation

For implementing this idea, we are choosing a dataset from the COVID-19 period. In this dataset, we have the number of deaths due to the COVID-19 infection in a certain period of time, in each country. Based on that number, we can use data visualisation techniques to understand the trends, to see it's a pandemic. On a small-scale implementation, we can use local records(if available) to determine whether a certain infection has the potential to become an epidemic.

Here are the datasets used for this purpose:

1. https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv
2. https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_deaths_global.csv
3. https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_daily_reports/01-15-2023.csv

Using data visualization techniques and Python libraries, we can see the following results:

The following libraries are imported:

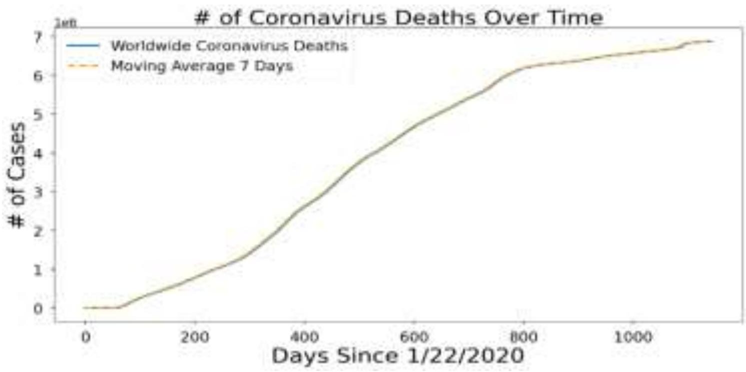
CODE:

```
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.colors as mcolors
import pandas as pd
import random
import math
import time
from sklearn.linear_model import LinearRegression,
BayesianRidge
from sklearn.model_selection import RandomizedSearchCV,
train_test_split
from sklearn.preprocessing import PolynomialFeatures
from sklearn.svm import SVR
from sklearn.metrics import mean_squared_error,
mean_absolute_error
import datetime
import operator
%matplotlib inline
import warnings
plt.style.use('seaborn-poster')
```

CODE:

```
adjusted_dates = adjusted_dates.reshape(1, -1)[0]
plt.figure(figsize=(16, 8))
plt.plot(adjusted_dates, total_deaths)
plt.plot(adjusted_dates, world_death_avg, linestyle='dashed',
color='orange')
plt.title('# of Coronavirus Deaths Over Time', size=30)
plt.xlabel('Days Since 1/22/2020', size=30)
plt.ylabel('# of Cases', size=30)
plt.legend(['Worldwide Coronavirus Deaths', 'Moving Average {}
Days'.format(window)], prop={'size': 20})
plt.xticks(size=20)
plt.yticks(size=20)
plt.show()
```

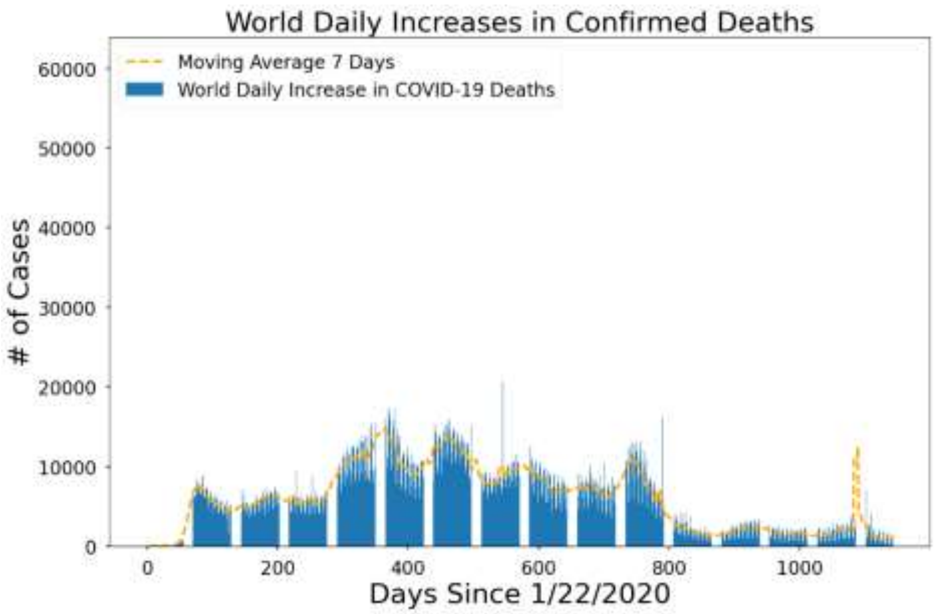
OUTPUT:



CODE:

```
plt.figure(figsize=(16, 10))
plt.bar(adjusted_dates, world_daily_increase)
plt.plot(adjusted_dates, world_daily_increase_avg,
color='orange', linestyle='dashed')
plt.title('World Daily Increases in Confirmed Cases', size=30)
plt.xlabel('Days Since 1/22/2020', size=30)
plt.ylabel('# of Cases', size=30)
plt.legend(['Moving Average {} Days'.format(window), 'World
Daily Increase in COVID-19 Cases'], prop={'size': 20})
plt.xticks(size=20)
plt.yticks(size=20)
plt.show()
```

OUTPUT:



Now, we define the following functions to better visualize our data, to know the condition of each country.

CODE:

```
def country_plot(x, y1, y2, y3, country):
    confirmed_avg = moving_average(y1, window)
    confirmed_increase_avg = moving_average(y2, window)
    death_increase_avg = moving_average(y3, window)
```

```
    SIZE = (12, 8)


    plt.figure(figsize=SIZE)
    plt.plot(x, y1)
    plt.plot(x, confirmed_avg, color='red', linestyle='dashed')
    plt.legend(['{} Confirmed Cases'.format(country), 'Moving
Average {} Days'.format(window)], prop={'size': 20})
    plt.title('{} Confirmed Cases'.format(country), size=30)
    plt.xlabel('Days Since 1/22/2020', size=30)
  plt.ylabel('# of Cases', size=30)
    plt.xticks(size=20)
    plt.yticks(size=20)
    plt.show()


    plt.figure(figsize=SIZE)
    plt.bar(x, y2)
    plt.plot(x, confirmed_increase_avg, color='red',
linestyle='dashed')
    plt.legend(['Moving Average {} Days'.format(window), '{} Daily
Increase in Confirmed Cases'.format(country)], prop={'size': 20})
    plt.title('{} Daily Increases in Confirmed
Cases'.format(country), size=30)
    plt.xlabel('Days Since 1/22/2020', size=30)
    plt.ylabel('# of Cases', size=30)
    plt.xticks(size=20)
    plt.yticks(size=20)
    plt.show()


    plt.figure(figsize=SIZE)
    plt.bar(x, y3)
    plt.plot(x, death_increase_avg, color='red', linestyle='dashed')
    plt.legend(['Moving Average {} Days'.format(window), '{} Daily
Increase in Confirmed Deaths'.format(country)], prop={'size': 20})
    plt.title('{} Daily Increases in Deaths'.format(country),
size=30)
```

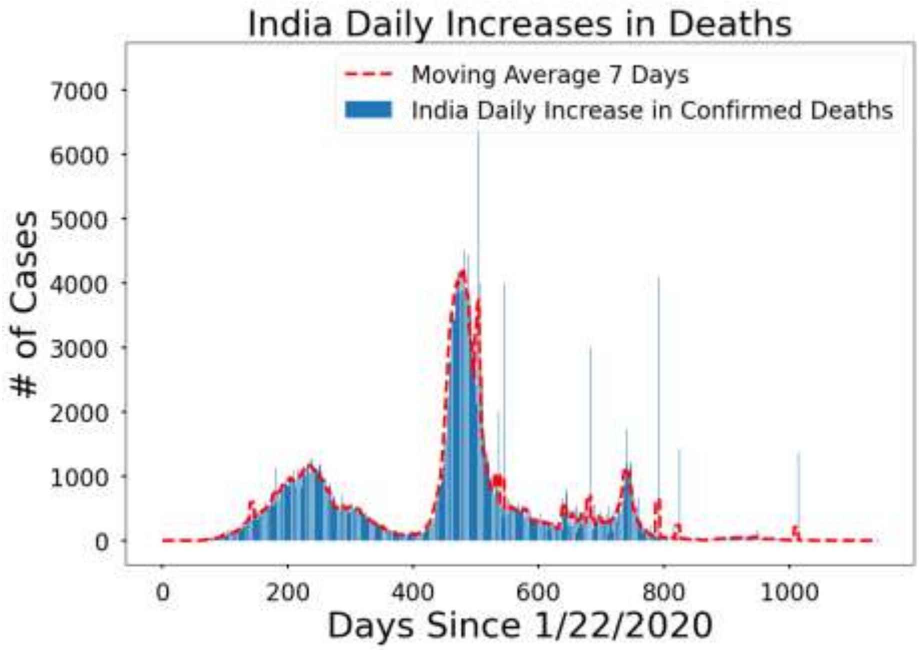
```
    plt.xlabel('Days Since 1/22/2020', size=30)
    plt.ylabel('# of Cases', size=30)
    plt.xticks(size=20)
    plt.yticks(size=20)
    plt.show()



# helper function for getting a country's total covid cases and 
deaths
def get_country_info(country_name):
    country_cases = []
    country_deaths = []

    for i in range(num_dates):

country_cases.append(confirmed_df[confirmed_df['Country/Region']==co
untry_name][ck[i]].sum())

country_deaths.append(deaths_df[deaths_df['Country/Region']==country
_name][dk[i]].sum())
    return (country_cases, country_deaths)



def country_visualizations(country_name):
    country_info = get_country_info(country_name)
    country_cases = country_info[0]
    country_deaths = country_info[1]

    country_daily_increase = daily_increase(country_cases)
    country_daily_death = daily_increase(country_deaths)

    country_plot(adjusted_dates, country_cases, 
country_daily_increase, country_daily_death, country_name)
```

Now, we call this country_visualizations(country_name) function for each country.

CODE:

```
countries = ['India']
for country in countries:
    country_visualizations(country)
```

OUTPUT:

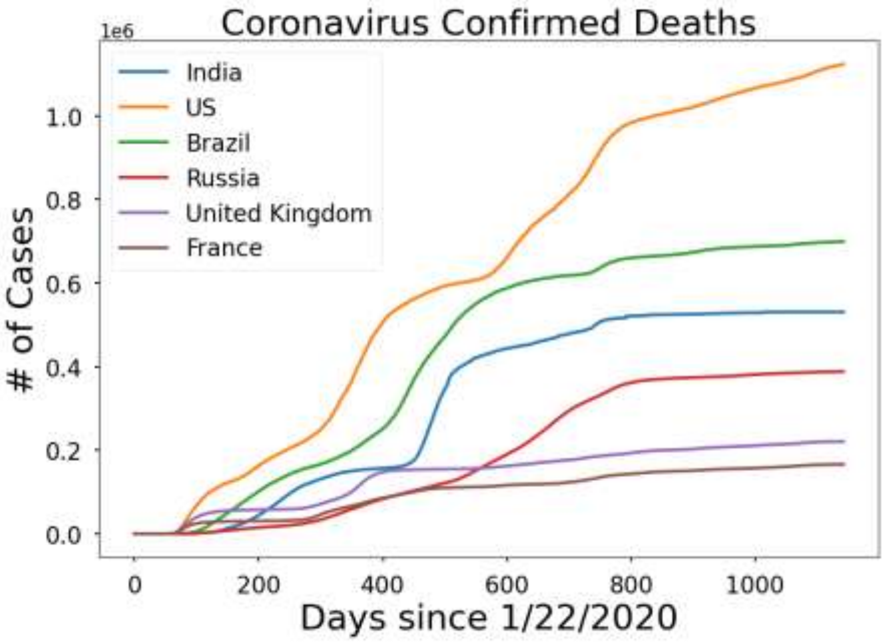


CODE:

```
# Country Comparison
compare_countries = ['India', 'US', 'Brazil', 'Russia', 'United Kingdom', 'France']
graph_name = ['Coronavirus Confirmed Cases', 'Coronavirus Confirmed Deaths']

for num in range(2):
    plt.figure(figsize=(12, 8))
    for country in compare_countries:
        plt.plot(get_country_info(country)[num])
```

```
plt.legend(compare_countries, prop={'size': 20})
plt.xlabel('Days since 1/22/2020', size=30)
plt.ylabel('# of Cases', size=30)
plt.title(graph_name[num], size=30)
plt.xticks(size=20)
plt.yticks(size=20)
plt.show()
```

OUTPUT:



This is a Python code for visualizing and comparing COVID-19 data using various machine learning and data visualization techniques. Let's break down the key components:

1. **Data Visualization:**
    - The script uses the **matplotlib** library to create time-series plots and bar charts for COVID-19 data.
    - The charts include:
        - Worldwide COVID-19 cases and deaths over time.
        - Daily increases in confirmed cases and deaths globally.
        - Country-specific plots for confirmed cases, daily increases, and daily deaths.
2. **Data Processing and Modeling:**

- The script imports data manipulation libraries such as **numpy** and **pandas** for handling and processing the COVID-19 data.
- Machine learning algorithms from **scikit-learn** can be utilized, including:
    - **Linear Regression** and **BayesianRidge** for trend analysis.
    - **PolynomialFeatures** for feature engineering.
    - **SVR (Support Vector Regression)** for modeling.

3. **Country-Specific Visualizations:**
    - Functions are defined to generate visualizations for individual countries, showcasing confirmed cases, daily increases, and daily deaths.
4. **Comparison of Multiple Countries:**
    - The script provides a comparison of COVID-19 cases and deaths among selected countries, allowing for an overview of how different regions are affected.
5. **Epidemic Detection/Prediction:**
    - The script's design and functionality can be useful for analyzing the progression of the COVID-19 pandemic and identifying patterns or trends.
    - The machine learning algorithms, such as SVR and linear regression, could potentially be used for making predictions about future case numbers.

How this Model Can be Used for Epidemic Detection/Prediction:

- The machine learning algorithms can learn patterns from historical data to make predictions about future trends in COVID-19 cases and deaths.
- By analyzing the visualizations and using the algorithms, one can identify potential hotspots, estimate the impact of interventions, and predict the trajectory of the epidemic in different regions.
- The model's effectiveness depends on the quality and representativeness of the input data and the appropriateness of the chosen algorithms.

Hence, we can see that the idea was to integrate data visualization and machine learning techniques to analyze and compare COVID-19 data globally and on a country-specific level. While it provides valuable insights, users should interpret results cautiously and consider the dynamic nature of the pandemic. The model could be enhanced by incorporating more advanced forecasting methods and considering additional relevant features.

Applicable Patents

In the context of epidemic detection specific to India, potential patent considerations encompass machine learning algorithms tailored for the Indian context, advancements in algorithmic methodologies, and model training techniques. Additionally, patents related to data integration methods, real-time data processing, and analytics platforms designed for epidemic data analysis in

India can be explored. Further information about patents covering predictive analytics methods adjusted for forecasting epidemics, geospatial analysis, user interface design, and data integration techniques are recommended. Additionally, patents associated with telemedicine and remote monitoring technologies, considering unique healthcare challenges in India, are important for comprehensive coverage in epidemic detection.

Applicable Regulations

The project necessitates compliance with regulations, including:

- Adherence to HIPAA and GDPR for safeguarding patient health information and ensuring data protection.
- Ethical approval and consent for handling sensitive healthcare data beyond legal requirements.
- Compliance with data sharing agreements and regulations when collaborating with healthcare institutions.
- Awareness of environmental regulations, particularly if the model incorporates environmental and climate data.
- Implementation of robust data security measures, including encryption and access controls, to prevent data breaches.
- Providing transparency in model operations and being accountable for predictions to align with regulatory requirements.
- Adherence to ethical AI principles outlined by organizations like the IEEE or ACM for ensuring fairness, transparency, and accountability.
- Consideration of environmental impact assessments if the model significantly incorporates environmental data.
- Compliance with agreements and regulations related to healthcare and environmental data for international operation.
- Ensuring accessibility compliance for user interfaces and outputs to cater to individuals with disabilities.

Strict adherence to these regulations is critical for ethical and legal data utilization, minimizing potential legal complications and instilling confidence in the project.

Applicable Constraints

- **Budget Constraints:** Limited funding can impact project scope and resources.
- **Data Quality and Availability:** Inadequate data quality and availability can hinder model accuracy.
- **Expertise and Talent:** A shortage of skilled professionals can affect project execution.

- **Regulatory Compliance:** Meeting data privacy and ethical regulations is critical.
- **Time Sensitivity:** Timely model development and deployment are essential for epidemic response.
- **Infrastructure and Technology:** Adequate computing resources and technology are necessary.
- **Geographic Coverage:** Data coverage constraints affect the model's applicability.
- **Scalability:** The model should scale to handle increased data and demand.
- **Community Acceptance:** Trust and acceptance within the community are vital.
- **Interoperability:** Integration with existing systems can be challenging.

Balancing these constraints is key to the success of an epidemic prediction model.

Business Model (Monetization Idea)

This epidemic prediction application can be very useful and beneficial for business too. Healthcare is one of the most essential and important aspects for any country or region. We can hence, monetize this model by sharing the data with data analysts, medical teams, epidemiologists, social activism campaigns, etc.

- **Data Sharing with Stakeholders:**

  Sharing the model's predictions with data analysts, medical teams, epidemiologists, and social activism campaigns can be monetized through a commission-based model. For instance, campaigns utilizing the data could yield a commission of 5%.

- **Licensing Fees for Research Organizations:**

  Organizations seeking to use the model's data for research or services can be charged licensing fees. A 10% licensing fee could be negotiated, contributing to revenue generation.

- **API Access for Applications and Platforms:**

  Charging other applications and platforms a fee of 5% for accessing epidemic prediction data via APIs designed with this model can be a revenue stream. This monetization model leverages the accessibility and accuracy of the prediction data.

- **Sale to Government Agencies:**

  Selling the prediction tool to government agencies presents an opportunity for negotiation. A commission of 5% on the contract value could be established, contributing to profitability.

These monetization strategies align with the project's goals, providing a sustainable business model while ensuring that valuable predictions contribute to various sectors. It's essential to establish clear terms and agreements, emphasizing the ethical use of data and compliance with relevant regulations. Additionally, ongoing collaboration and refinement of the prediction model can enhance its value and marketability.

Concept Generation

The idea for this epidemic detection model stems from the profound impact of COVID-19 worldwide. Recognizing the need for proactive measures, the concept revolves around predicting and preventing outbreaks before they escalate into significant threats. By harnessing local health data and employing machine learning algorithms, the goal is to identify potential epidemics in their early stages, considering factors such as causative agents, transmission rates, environmental triggers, and more. The aim is to intervene promptly, preventing localized issues from evolving into global pandemics, similar to the trajectory of the COVID-19 virus.

1. **Early-Stage Detection:**

   Using machine learning algorithms to analyze local health data and detect potential epidemics in their initial phases.

2. **Predictive Factors:**

   Considering a comprehensive set of factors, including the cause of the outbreak, transmission rates, and environmental triggers, to enhance the predictive capabilities of the model.

3. **Preventive Measures:**

   Developing a system that not only predicts potential epidemics but also suggests preventive measures. This could include recommendations for containment strategies, resource allocation, and targeted interventions.

4. **Global Impact Mitigation:**

   Focusing on preventing the escalation of local health issues to global proportions, aligning with the broader goal of curbing the spread of infectious diseases.

5. **Real-Time Monitoring:**

   Implementing real-time monitoring capabilities to ensure timely updates and interventions, facilitating swift responses to emerging health threats.

6. **Interdisciplinary Collaboration:**

   Fostering collaboration between data scientists, healthcare professionals, and policymakers to integrate diverse expertise, ensuring the model's effectiveness and alignment with practical health interventions.

7. **Continuous Improvement:**

   Establish mechanisms for continuous refinement of the model based on evolving health data, emerging threats, and the efficacy of preventive measures.

Concept Development

In developing our epidemic prediction model, we leverage diverse datasets, including localized health data, historical epidemic records, current spread information, healthcare details, environmental data, international travel insights, vaccination coverage, and healthcare infrastructure statistics. Utilizing sample data from the COVID-19 pandemic as a starting point, the model aims to produce actionable outputs such as early warnings, disease spread predictions, and intervention recommendations. By foreseeing health crises, the app facilitates proactive measures, safeguarding public health.

Final Report Prototype

Front-End:

1. **Dashboard:** A user-friendly dashboard offers intuitive visualizations for insights into epidemic data.
2. **User-Friendly Interface:** The interface ensures ease of use, allowing users to explore and analyze data effortlessly.
3. **Dynamic Charts and Design:** Dynamic charts provide real-time updates, and the design ensures seamless usability across devices.
4. **Clear Summaries:** Key findings are presented through clear summaries, enabling quick comprehension of critical information.

Back-End:

1. **Data Processing:** Data processing includes handling missing values, ensuring data integrity for accurate predictions.
2. **Machine Learning Model Integration:** The integration of a powerful Gradient Boosting model enhances the model's predictive capabilities.
3. **Real-time Database Support:** Database support facilitates real-time updates, ensuring the model adapts swiftly to dynamic epidemic data.
4. **API Integration and Secure Data Handling:** Seamless API integration enables efficient data sharing, and robust security measures ensure safe handling of sensitive information.
5. **Scalable Architecture:** The model's architecture is designed for scalability, accommodating increased data volumes and evolving requirements.
6. **Performance Optimization:** Continuous optimization measures are implemented for enhanced system performance, ensuring efficiency in processing and analysis.
7. **Logging and Monitoring:** Logging and monitoring mechanisms are in place to track system health, providing insights into operational status and facilitating timely interventions.

This prototype outlines a comprehensive and user-centric approach for both the front-end and back-end aspects of the epidemic detection model. The design prioritizes accessibility, real-time functionality, security, and scalability to create a robust and effective tool for proactive health management.

Product Overview

Utilizing established technologies, such as Python, ensure technical viability. The user-friendly interface and scalable backend enhance operational efficiency, adhering to data protection laws for legal compliance.

In terms of viability, the model addresses the demand for predictive analytics in healthcare, balancing initial investment with potential revenue streams.

For monetization, the model adopts a subscription model with tiered plans, collaborates for enterprise-level licensing with customized solutions, and charges fees for API access, generating revenue through integration.

The model utilizes data analysis and predictive algorithms for early epidemic detection, drawing from diverse sources like historical epidemic data, current healthcare information, and more. Machine learning algorithms, real-time data integration, and cloud computing contribute to its functionality.

The multidisciplinary team involves epidemiologists, healthcare experts, data scientists, software developers, UX/UI designers, and project managers.

Cost factors include algorithm complexity, data sources, and ongoing operational expenses for data acquisition and model maintenance.

Business Modeling

The business model for Epidemic Detection focuses on delivering proactive and accurate insights into potential health crises, benefiting both healthcare authorities and the general public. Through advanced predictive analytics, the platform empowers authorities with data-driven strategies for timely interventions, while individuals gain access to early warnings and recommended precautions.

1. **Revenue Generation:** The platform adopts a subscription-based model, providing healthcare authorities and institutions with advanced features for epidemic monitoring. Additionally, fees are charged for API access, facilitating integration into existing healthcare systems and generating revenue.
2. **Key Activities:** Continuous refinement of the predictive model and user engagement are key activities, supported by a multidisciplinary team comprising epidemiologists, healthcare experts, data scientists, software developers, and UX/UI designers.
3. **User Interaction:** A user-friendly interface ensures positive user experiences, facilitating effective communication of epidemic insights. Customer support plays a vital role in fostering strong relationships with healthcare authorities and institutions.
4. **Market Expansion:** Strategic partnerships with healthcare institutions, agencies, and online platforms enhance market reach, enabling the platform to serve a broader audience and contribute more effectively to public health management.
5. **Overall Objective:** The business model aims to provide a valuable and efficient solution in the dynamic landscape of epidemic detection, prioritizing timely interventions and public well-being.

Financial Equation

Here’s the financial model equation for the Epidemic Detection Model:

Financial Equation:

$Y = m \times x(t) - C$

Given Parameters:

- m: Subscription model price for the epidemic detection platform.
- $x(t)$: Total sales as a function of time.
- C: Total production and service cost.

Let's Choose Optimal Values:

- $m = Rs.300$ (Optimal subscription model price)
- $x(t) = Rs.8Cr$ (Estimated total sales as a function of time)
- $C = Rs.1.5L$ (Optimal total production and service cost)

Financial Equation for our model:

$$Y = 300 \times 8Cr - 1.5L$$

This financial model provides an approximate estimation of the total profit generated by the Epidemic Detection Model based on the chosen variables. It considers the subscription model price, total sales over time, and the associated production and service costs.

Conclusion

In conclusion, our epidemic prediction app represents a significant step forward in safeguarding global health. Inspired by the lessons of the COVID-19 pandemic, this app harnesses the power of data analytics, machine learning, and real-time information to provide early warnings of potential epidemic outbreaks. By analysing diverse datasets and patterns, it empowers healthcare authorities and professionals to take proactive measures and mitigate the impact of diseases.

GitHub link:
https://github.com/AparajitaGoswami/Feynn_Labs/blob/7d3f7a665dbcb0abbd54c906af3c452734dc2e3c/Project_3/epidemicpredict.ipynb