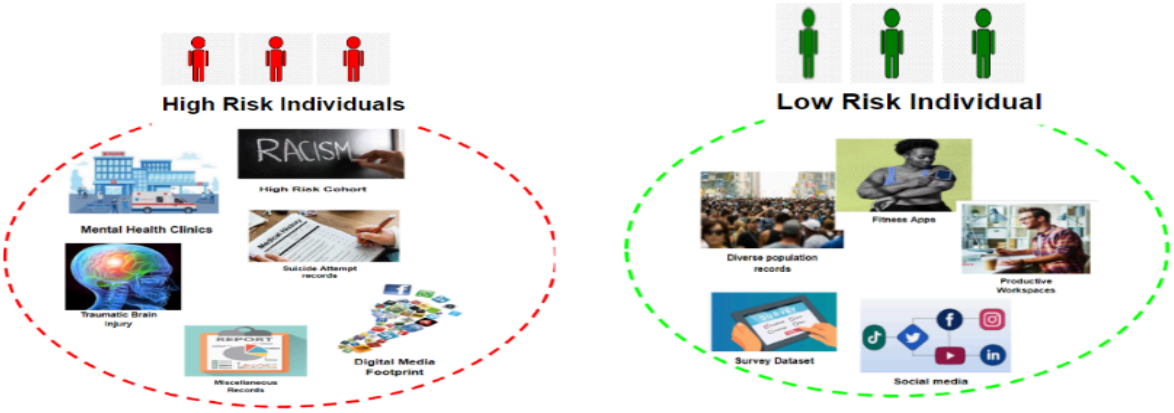
Problem Definition:One Pager

AI Suicide Intervention:One Pager

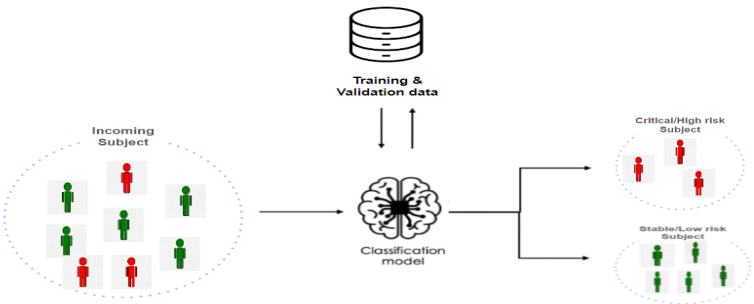
1. Identify High/Low Risk individuals



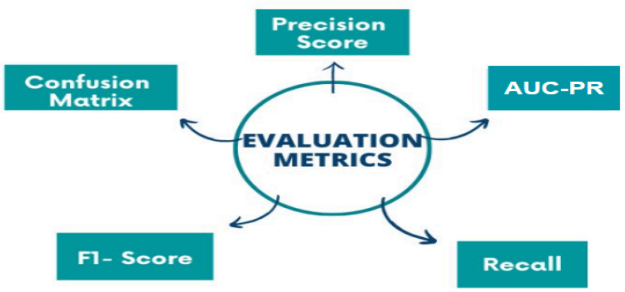
2. Prepare Data Records

Locations Info	Browser History	Communication	Health Data	Finances	App Usage	Photo Clicks	Social Media History	Risk
Visits Isolated places	Searches like "how to stop feeling depressed"	whatsapp chat with friends like "Feeling Low"	erractic sleep patterns	overdue payments on utility bills	frequent visits on adult/social media sites	Sunsets/Empty streets etc.	Posts like "I feel like I am not needed anymore"	
Hiking, community/social events	Gratitude, positive productive searches	Always positive, optimistic communication	10k walking steps	Maintains diverse healthy portfolio with low debt	Educational websites	Family/Friend/Out door activities	Shares positive life experiences and engage in positive productive posts.	

3. Train Classification Machine Learning Model



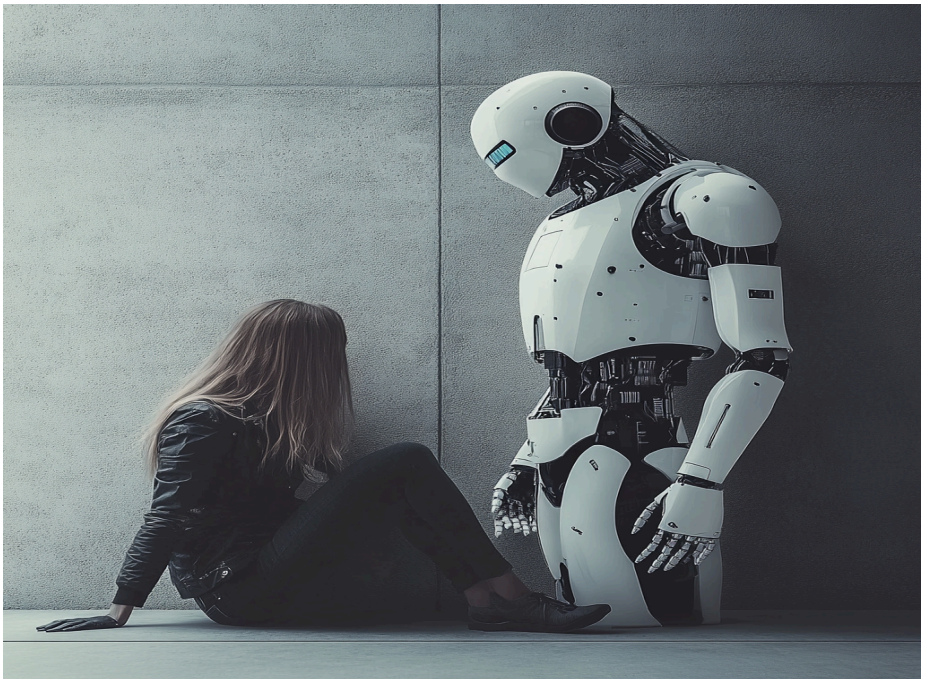
4. Report Metrics:Accuracy is misleading metric in unbalanced problem statement like suicide prediction.



Problem Definition:Detailed Document

Problem statement

The AI Suicide Intervention Project seeks to harness the power of artificial intelligence to predict and prevent suicidal tendencies, offering a revolutionary approach to mental health care. Suicide is a devastating event that leaves families and communities grappling with loss and searching for answers. This project aims to provide those answers by developing an AI model capable of identifying individuals at risk of suicide, thereby enabling timely and effective interventions.



In today's digital age, individuals generate vast amounts of data through their interactions with smartphones, computers, and other connected devices. These digital traces can provide crucial insights into a person's mental state and potential risk factors for suicide. Our goal is to create a sophisticated AI model that will be trained on dataset that includes a wide array of digital footprints and behavioral data that can predict suicidal behavior with unprecedented accuracy. Each of these data types offers unique insights into an individual's behavior, emotional state, and social interactions, which are critical for identifying early warning signs of suicidal tendencies.

The AI model will be designed to analyse large datasets and detect patterns and relationships that may be invisible to human observers. These insights will serve a dual purpose:

1. advancing research in suicide prevention by identifying complex behavioral patterns associated with suicidal tendencies, and
2. providing a software solution that can assess an individual's risk of suicide given sufficient data and thereby facilitates prevention through appropriate intervention..

Although privacy concerns present challenges for widespread deployment, our primary and critical goal remains to save lives without compromising user privacy.

Data Collection

The effectiveness of a suicide prediction model depends on the quality and diversity of the data used during training. To build a robust model capable of accurately distinguishing between at-risk individuals (positive samples) and those not at risk (negative samples), it is essential to gather comprehensive and representative datasets for both classes.

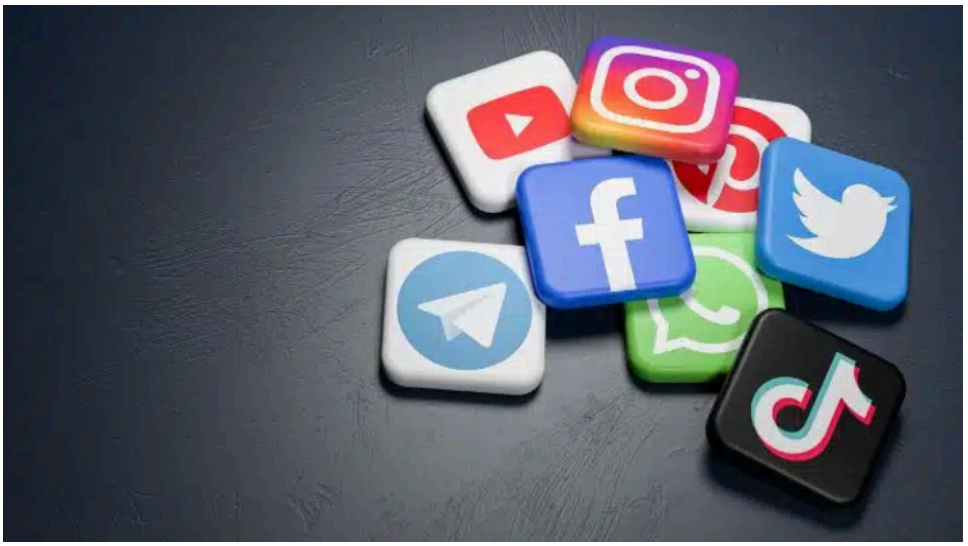
Data Collection for Positives ("high risk" individuals)

Positive samples represent individuals who are at high risk of suicide or have already engaged in suicidal behavior. Collecting data for this group requires a sensitive, ethical, and multi-faceted approach.

1. **Clinical Data Sources**:
    - **Mental Health Clinics:** Data from individuals who have been diagnosed with severe depression, anxiety, or other mental health disorders associated with suicide risk can be sourced from psychiatric hospitals and clinics. This includes electronic health records (EHRs) containing diagnoses, treatment history, and psychological assessments.

    - **Suicide Attempt Records:** Information on individuals who have survived suicide attempts can be obtained from medical records and emergency services. This data provides critical insights into high-risk behaviours and triggers.
    - **Traumatic Brain Injury**: Collecting data from patients with traumatic brain injury (TBI) for a suicide prediction model is critically important due to the elevated risk of suicidal thoughts and behaviors within this population. TBI often results in significant cognitive, emotional, and behavioral changes that can increase vulnerability to mental health issues, including depression and suicidal ideation. By incorporating data from TBI patients, the model can gain deeper insights into how brain injuries affect mental health and contribute to suicide risk. This understanding is crucial for developing targeted interventions and preventive strategies tailored to the unique needs of individuals with TBI.

2. **Forensic Analysis of Digital Footprints**
    - **Posthumous Data Collection:** In cases where an individual has died by suicide, with consent from grieving families, forensic analysis of the deceased's digital footprint can be conducted. This involves gathering data from the individual's smartphones, computers, and other digital devices, including location history, browsing activity, communication logs, and social media interactions.



    - **Social Media and Online Forums:** Data from social media platforms and suicide-related online forums where individuals express suicidal thoughts or seek help can be valuable. This includes analysing text posts, comments, and shared multimedia content.
3. **Longitudinal Studies:**
    - **Follow-up Studies:** Longitudinal studies that track individuals over time, particularly those with known risk factors such as a history of mental illness or previous suicide attempts, can provide data on changes in behavior, mood, and social interactions leading up to a suicide attempt or completion.

- **High-Risk Cohorts:** Identifying and monitoring high-risk groups, such as individuals undergoing significant life stressors (e.g., job loss, divorce, gender bias, racism) or those with a family history of suicide, can help in gathering time-sensitive data that captures the escalation of suicidal tendencies.

4. **Anonymous data collection:** Partnering with non-profit organizations that support suicide prevention can encourage individuals to anonymously donate their data for research. This can include data from mental health apps, wearable devices, and online mental health platforms.

Data Collection for Negatives ("no risk" individuals)

Negative samples consist of individuals who are not at risk of suicide and serve as the control group. Collecting this data requires a broad and representative approach:

1. **General Population Data:** EHRs from general practitioners that indicate the absence of mental health disorders or suicide risk factors can serve as a source for negative samples. This includes individuals who have routine check-ups but no significant mental health concerns. To ensure the model generalises well across different populations, it's important to include data from individuals of various ages, genders, ethnicities, socioeconomic backgrounds, and geographic locations.
2. **Behavioral Data from Low-Risk Groups:**
    - **Fitness and Health App Data:** Data from fitness trackers and health apps from individuals leading a healthy lifestyle with no indication of mental distress can provide a baseline for normal behavioural patterns.

    - **Social Media Usage:** Analysing social media behaviour of users who engage positively online, maintain social connections, and exhibit no signs of distress can help in understanding typical digital interactions.
3. **Random Sampling from Large Datasets**
    - **Census and Survey Data:** Randomly sampling individuals from large-scale datasets like national censuses or health surveys can provide a

representative control group. This data should be pre-screened to exclude individuals with known risk factors.

- **Workplace and Educational Records:** Data from workplaces or educational institutions that track productivity, attendance, and social engagement can be used to represent individuals not at risk, especially when there are no signs of mental health issues.

4. **Longitudinal Studies for tracking stable individuals:**Similar to longitudinal studies for at-risk groups, tracking individuals who have shown no signs of mental health deterioration over time can help create a control dataset. This group should include individuals who have maintained stable employment, relationships, and social activities.

Comprehensive Data Collection for Suicide Prediction: Types, Examples, and Analysis

To build a robust AI suicide prediction model, collecting and analyzing various types of data can provide valuable insights into the behavioral patterns and emotional states of individuals at risk of suicide. Below, each data type is expanded with detailed examples to illustrate what data from a suicide victim might look like and how it can be used to train a suicide prediction model.

1. Location Data: GPS Information Revealing Patterns of Movement and Isolation

Example:
Imagine a young adult who was once socially active, frequently visiting friends, attending social events, and participating in outdoor activities. Over several months, their GPS data reveals a significant change in behaviour. The individual's movement patterns become increasingly restricted to their home, with only occasional trips to places of necessity, such

as grocery stores or work. There are no records of visiting friends or engaging in previous hobbies, like hiking or attending community events. The data may also show prolonged stays at locations associated with solitude or introspection, such as parks or cemeteries.

Analysis Use:
This data can indicate social withdrawal and isolation, which are key risk factors for suicide. The model can learn to recognize these patterns as potential indicators of increased suicide risk, particularly when combined with other data points like communication logs or health data.

2. Browsing History: Websites Visited, Search Queries, and Online Behavior

Example:
A middle-aged individual who has been experiencing financial difficulties might start visiting websites related to financial aid, debt relief, or even online payday loans. Over time, their search queries shift towards topics related to mental health, depression, and eventually, suicide. Specific searches might include “how to stop feeling depressed,” “signs of severe depression,” and “how to commit suicide.” The individual may also spend time on forums or discussion boards where users talk about their struggles with mental health and share experiences related to suicidal thoughts.

Analysis Use:
This browsing history can be a critical indicator of the individual's mental state. The shift in online behavior from financial stress to suicidal ideation is a clear red flag that the AI model can learn to identify as a precursor to suicide risk.

3. Communication Logs: Emails, Text Messages, and Social Media Interactions to Detect Signs of Distress or Withdrawal

Example:
Consider a young professional who, over the past year, has increasingly sent text messages to close friends expressing feelings of hopelessness and despair. Early messages might have been subtle, such as “I’m just feeling down today,” but later evolved into more explicit statements like “I don’t know how much longer I can keep doing this” or “It would be better if I weren’t here.” In emails, they might start neglecting work responsibilities, missing deadlines, and responding with terse or vague answers. Their social media interactions also decline; they stop liking or commenting on posts and eventually reduce their overall online presence.

Analysis Use:
The AI model can analyse the tone and content of these communications to detect escalating distress. Patterns of reduced communication, along with increasingly negative language, are strong indicators that the individual is at risk of suicide. Sentiment analysis tools can be employed to quantify the emotional content of these messages.

4. Health Data: Information from Fitness Trackers and Health Apps, Such as Sleep Patterns, Physical Activity, and Heart Rate Variability

Example:
A person in their early 30s who was once physically active starts showing significant changes in their health data. Their fitness tracker reveals a drastic decrease in daily physical activity—steps per day drop from 10,000 to less than 2,000. Sleep patterns become erratic, with frequent awakenings, reduced overall sleep time, and long periods of inactivity during the day, suggesting daytime naps or lethargy. Heart rate variability decreases, which can be a sign of stress and poor mental health. Additionally, the individual might stop using the fitness tracker consistently, indicating a loss of interest in personal health. Medical records, such as prescription history, test results, imaging, blood work, and other pertinent information, offer additional context and can help clarify any underlying medical conditions or treatments that may be influencing these changes.

Analysis Use:
These changes in health data are critical indicators of deteriorating mental health. The AI model can be trained to recognize such patterns as warning signs, particularly when combined with other behavioral data, signaling an increased risk of suicide.

5. Financial Records: Bank Statements, Credit Card Transactions, and Spending Habits That May Indicate Financial Stress

Example:
A person in their late 40s facing long-term unemployment might exhibit changes in their financial behaviour. Bank statements reveal an increasing reliance on credit, with frequent small withdrawals that indicate financial strain. Over time, there are more declined transactions and overdue payments on utility bills. The individual might also start making uncharacteristic purchases, such as large sums spent on alcohol, gambling, or purchases related to suicide means (e.g., purchasing medications in large quantities).

Analysis Use:
Financial stress is a significant contributor to suicidal ideation. The AI model can use these financial records to identify individuals under financial duress, especially when correlated with other stress indicators, such as changes in communication patterns or health data.

6. App Usage: Frequency and Nature of App Interactions, Highlighting Changes in Routine or Interests

Example:
A university student who once actively used educational apps, social media, and messaging platforms starts showing a shift in app usage. They begin spending less time on educational apps and more on entertainment or distraction apps like gaming or streaming services. Over time, their interaction with productivity apps declines sharply, and they may even uninstall apps related to social interaction or learning. The data may also show a spike in the usage of mental health apps or apps related to crisis intervention, but with no subsequent improvement in app usage patterns.

Analysis Use:
The AI model can interpret these shifts in app usage as indicators of a change in mental state. A decline in productive app usage combined with increased engagement in passive or escapist activities can signal withdrawal and a decrease in cognitive engagement, both of which are risk factors for suicide.

7. Multimedia Files: Analysis of Photos, Videos, and Audio Recordings for Signs of Emotional Distress

Example:
An adolescent who was previously active in photography and social media might start posting fewer photos and, when they do, the content might shift in tone. Earlier photos might have been of friends, family, and outdoor activities, but newer photos could reflect darker themes, such as sunsets, empty streets, or images with melancholic filters. If the individual records audio diaries or videos, there might be a noticeable change in tone—expressions of sadness, despair, or reflections on difficult life experiences.

Analysis Use:
Multimedia analysis can reveal shifts in mood and emotional state. The AI model can be trained to detect these changes, such as the use of darker filters, melancholic themes, or changes in voice tone in audio recordings, which can serve as early indicators of suicidal ideation.

8. Social Media History: Analysis of All Social Media Accounts for Clues That May Indicate Stress, Withdrawal, Isolation, Etc.

Example:
A young adult who was once very active on social media gradually reduces their activity. Initially, they may post updates and engage with friends regularly, but over time, their posts become less frequent and more negative in tone. Posts might express feelings of being overwhelmed or isolated, such as “I’m so tired of everything” or “I feel like I’m not needed anymore.” They might also unfollow or unfriend people, reducing their social network, and start following pages or joining groups that discuss topics related to mental health struggles or existential themes.

Analysis Use:
Social media history can provide a rich dataset for understanding an individual's mental state. The AI model can analyze the frequency, content, and tone of posts, as well as changes in social connections, to detect signs of stress, withdrawal, and isolation—all of which are key indicators of suicide risk.

Ethical guidelines for dataset preparation

Voluntary Participation: All data collection should involve informed consent, ensuring participants understand the purpose, risks, and benefits of their involvement.

Transparency: Clearly communicate how the data will be used, stored, and protected, emphasising the model's goal to save lives and advance mental health research.

Anonymization: All personal identifiers should be removed to protect individuals' identities. Data should be stored securely, with strict access controls and encryption.

Ethical Review: The data collection process should undergo ethical review by an Institutional Review Board (IRB) or equivalent body to ensure compliance with ethical standards and regulations.

Cultural Sensitivity: Be mindful of cultural differences in attitudes toward suicide and mental health. Tailor data collection methods to respect cultural norms and practices.

Avoiding Bias: Ensure that the data collection process does not perpetuate cultural biases or stereotypes. This includes training data collectors on cultural sensitivity and using diverse teams that reflect the populations being studied.

Collect Only What is Necessary: Adhere to the principle of data minimization by collecting only the data that is necessary to achieve the study's objectives. Avoid collecting excessive or irrelevant data that could increase the risk of privacy breaches or ethical concerns.

Support for Participants: Provide mental health resources and support for participants, particularly those contributing data related to suicidal behavior.

Dealing with unbalanced dataset

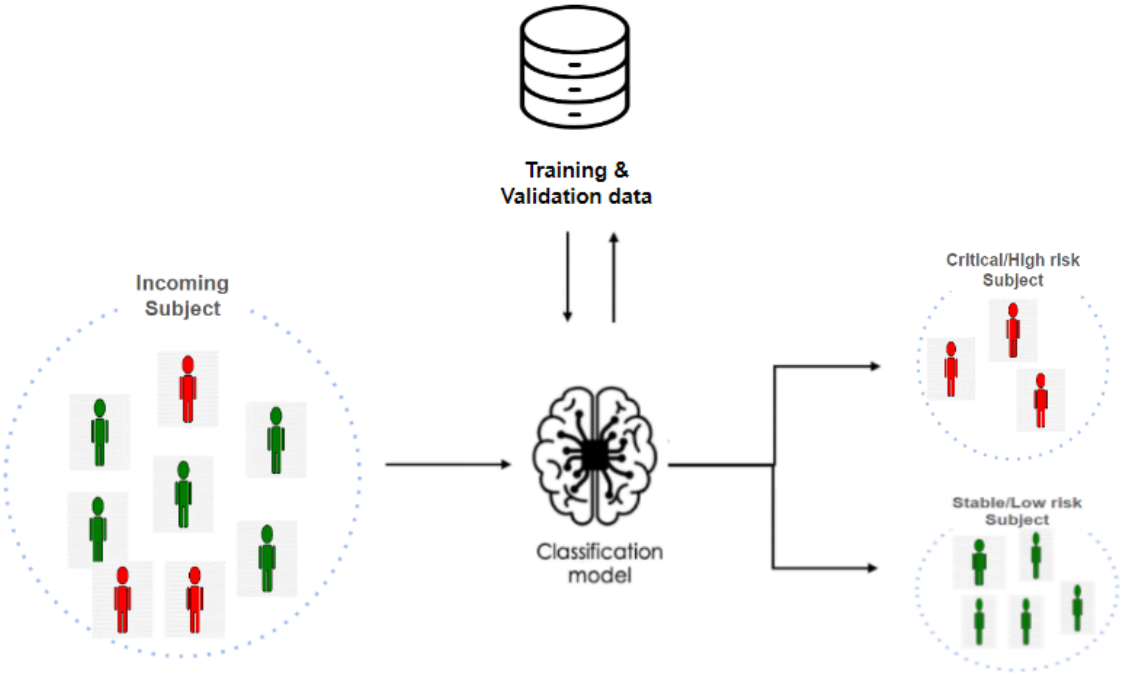
Given the rarity of suicide, the dataset is likely to be highly imbalanced, with far more negative samples than positive ones. In machine learning literature, there are various strategies to address unbalanced dataset e.g. Positive samples data augmentation, Synthetic Minority Over-sampling Technique (SMOTE), random undersampling and stratified sampling.

Machine Learning Formulation

We formulate the problem of AI suicide intervention as building a binary classification model. The objective of this model is to classify individuals into one of two categories: those at risk of suicide (positive class) and those not at risk (negative class). This binary classification task involves developing an algorithm that can process and analyze various types of data, ranging from digital footprints to personal health metrics, to make a prediction about an individual's suicide risk. The data utilised in this context is diverse and complex, encompassing aspects such as behavioural patterns, communication histories, health indicators, financial records and social interactions. Each of these data sources can potentially hold key insights into an individual's mental state, making the model's ability to integrate and analyse this information crucial to its success.

The problem formulation begins with the recognition that suicide is not a random event but rather the culmination of various factors that interact over time. These factors can include psychological issues such as depression or anxiety, social isolation, economic hardship, and

other stressors. Machine learning models are uniquely suited to identify the subtle, often nonlinear relationships between these factors, which might go unnoticed in traditional analysis methods. By treating the prediction of suicide as a binary classification problem, the model's task is to learn from past data where the outcomes are known—whether individuals have attempted or died by suicide or not—to make predictions about future cases.

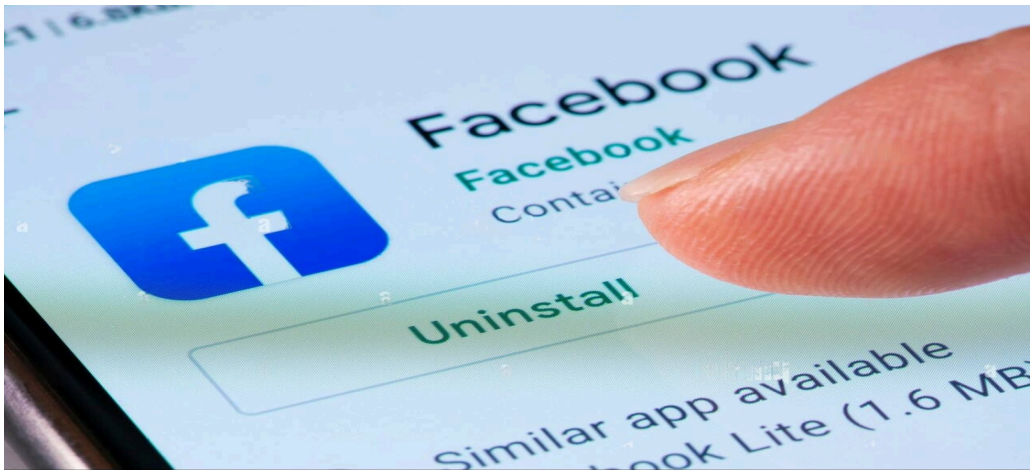


Given the sensitive nature of the problem, the model must be developed with a strong emphasis on ethical considerations. The implications of a model that incorrectly classifies someone as being at risk (false positive) or fails to identify someone who is at risk (false negative) are significant. A false positive could lead to unnecessary distress or intervention, while a false negative could mean a missed opportunity to prevent a tragedy. Therefore, the design of the model must prioritise minimising these errors, and the evaluation metrics must reflect the real-world impact of these predictions. The goal is to create a model that can be a powerful tool in the fight against suicide, providing early warnings and enabling timely interventions. **We'll address the user embedding learning part in detailed technical document**

From Suicide prediction to Churn Analysis: A comparative study of Binary classification tasks

1. **Churn Prediction:** Churn prediction is framed as a binary classification task where the goal is to categorise customers into two classes: those likely to churn (positive class) and those likely to stay (negative class). The model must analyse diverse data, such as customer interaction history, purchase behaviour, service usage patterns, and demographic information, to make predictions. Much like in suicide prediction, churn is influenced by multiple interacting factors, such as dissatisfaction with service, changes in personal circumstances, or external market conditions. Both scenarios require identifying subtle and often indirect indicators that might signal an

impending event. In churn prediction, this could be changes in engagement or satisfaction levels, while in suicide prediction, it might involve shifts in behavior or mental health indicators.



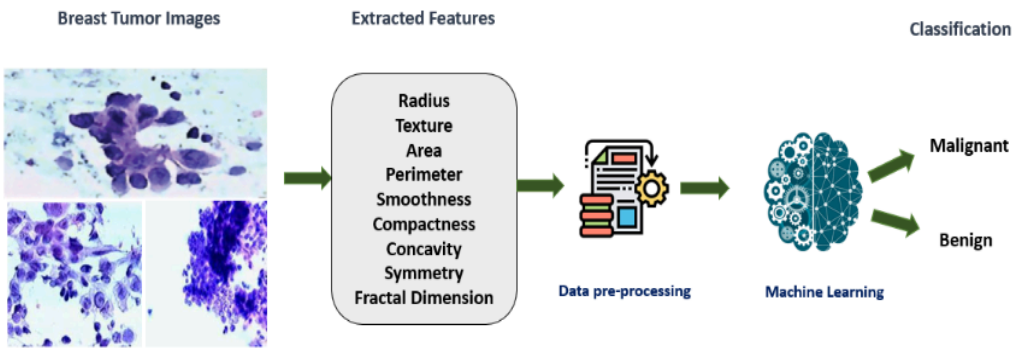
The challenge in both cases is to effectively handle an unbalanced dataset where the event of interest is relatively rare compared to non-events. This requires sophisticated modelling techniques to ensure that predictions are accurate and actionable.

2. **Fraud Detection**: In fraud detection, the problem is to build a binary classification model to identify fraudulent transactions (positive class) versus legitimate ones (negative class). The model leverages data such as transaction history, user behaviour patterns, and payment details to make predictions.

Similar to suicide prediction, fraud detection involves identifying unusual or suspicious patterns in the data, which may be subtle or complex. The task is particularly challenging due to the typically unbalanced nature of the dataset, where fraudulent transactions are much rarer than legitimate ones.

3. **Cancer(or Diabetes) Prediction**: Cancer prediction is also a binary classification problem where the aim is to classify patients as either having cancer (positive class)

or not having cancer (negative class). This model relies on medical data such as genetic information, imaging data, lifestyle factors, and family history to predict the likelihood of cancer.



Like suicide prediction, cancer prediction involves identifying non-linear relationships among various risk factors that contribute to the development of the disease. The data is often unbalanced, with far fewer cases of cancer compared to non-cancerous conditions, making accurate prediction crucial.

4. **Credit risk Assessment**: In credit risk assessment, the goal is to predict whether a borrower will default on a loan (positive class) or repay it (negative class). The model uses financial data such as credit history, income levels, employment status, and economic indicators to make this prediction. Financial behaviours such as missed payments, fluctuating income, or changes in employment status can signal financial distress.

Likewise, in suicide prediction, changes in behavior, social withdrawal, or prior mental health issues might suggest increased risk. Both tasks require integrating diverse data sources to detect subtle signals of an impending event. The challenge in both cases is managing imbalanced datasets, where the event of interest—default or suicide—is relatively rare compared to non-events. Thus, both problems necessitate advanced modelling techniques to accurately identify and mitigate risk despite the rarity of the adverse outcomes.

5. **Employee(or Student attrition) Attrition**: Employee attrition prediction involves classifying employees as likely to leave the company (positive class) or likely to stay (negative class). The model analyses data like job satisfaction, performance metrics, work environment, and compensation.

Like suicide prediction, employee attrition is often the result of various interacting factors, and the model's ability to discern these relationships is critical. Unbalanced datasets are common here as well, with fewer employees leaving than staying.

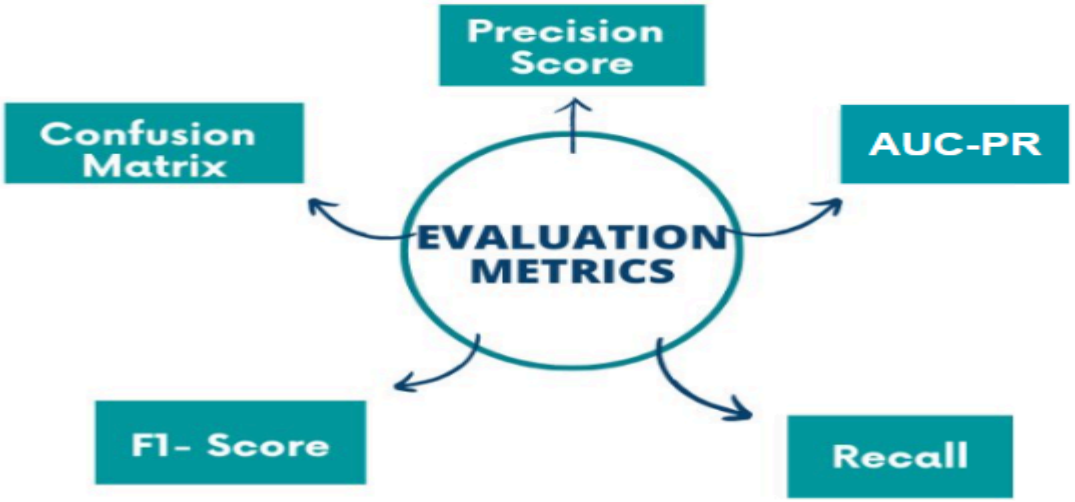
6. **Speech Recognition:** In speech recognition, the goal is to classify spoken language into specific commands or responses (positive class) versus non-command or irrelevant speech (negative class). The model processes audio features such as tone, pitch, cadence, and linguistic content to differentiate between meaningful commands and general speech. For example, virtual assistants like Alexa or Google Assistant must accurately recognize activation phrases like "Alexa" or "Hey Google" amidst everyday conversations and background noise.

Similar to suicide prediction, speech recognition involves handling complex, high-dimensional data where the target events (activation phrases) are relatively rare compared to non-target speech.

Evaluation Metrics

The suicide prediction model is a binary classification model, and the imbalanced nature of the dataset used in this problem leads to significant consequences for both false positives

and false negatives. Proper evaluation of the model is critical to ensure that it performs effectively in real-world scenarios, where the impact of errors can be profound. Hence Evaluating a suicide prediction model involves understanding the trade-offs each metric presents.



To effectively assess the model's performance, we use a train/test split. This means dividing the data into two parts: one for training the model and one for testing it. The **training data** is used to teach the model how to make predictions, while the **test data** is kept separate and unseen during training. By evaluating the model on this test data, we can check how well it performs on new, unseen data. This approach helps us ensure that the model doesn't just memorise the training data but can generalise and perform well with new information in practical situations. Let's discuss each metric, illustrating its pros and cons with a specific example related to suicide prediction.

1) Accuracy: Accuracy is the proportion of correct predictions out of the total number of cases.

- **Pros:** Easy to understand and compute.
- **Cons:** Can be misleading in imbalanced datasets.
- **Example**: Consider a dataset where 1,000 individuals are evaluated, and 10 of them are actually at risk of suicide (positive class). Now, suppose your model predicts that none of these individuals are at risk.In that case, Accuracy=990/1000=99%, i.e. a model that predicts "no risk" for everyone would still achieve 99% accuracy, though it fails to identify those at risk. Despite high accuracy, the model fails to identify the individuals at risk. This high accuracy does not reflect the poor performance in predicting the minority class.
- **Takeaway**: Accuracy alone is not reliable in cases with imbalanced data like suicide prediction.

2) Confusion Matrix: A confusion matrix is a table used to evaluate the performance of a classification model. It provides a detailed breakdown of the model's predictions by comparing them against the actual values. This matrix helps in understanding how well the model is performing in terms of different types of errors and correct classifications.

The confusion matrix typically has four key components:

	Predicted Positive	Predicted Negative
Actual Positive	TP (True Positive)	FN (False Negative)
Actual Negative	FP (False Positive)	TN (True Negative)

Structure of Confusion Matrix

For Example, let's say we have a dataset of 1000 individuals and 50 individuals are at "high-risk" i.e. total negatives=950 and our model's performance is the following:-

->Model predicted 15 positives and 985 as negatives.

->10 out of 15 positive predictions are correct and 945 negative predictions are correct.

Model's performance in confusion matrix format will look like this:-

	Predicted Positive	Predicted Negative
Actual Positive	10	40
Actual Negative	5	945

A. **True Positive (TP):** The number of instances where the model correctly predicted the positive class (e.g., correctly predicting an individual is at risk of suicide when they actually are at risk).It indicates how many actual positives are correctly identified by the model.

   Example: If an individual who is actually at risk of suicide is predicted to be at risk, this is a true positive.

B. **False Positive (FP):**The number of instances where the model incorrectly predicted the positive class (e.g., incorrectly predicting an individual is at risk of suicide when they are not at risk). It shows how many times the model mistakenly predicted the positive class.

   Example: If an individual who is not at risk of suicide is predicted to be at risk, this is a false positive.

C. **True Negative (TN):** The number of instances where the model correctly predicted the negative class (e.g., correctly predicting an individual is not at risk of suicide when they are not at risk). It reflects how many actual negatives are correctly identified by the model.

   Example: If an individual who is not at risk of suicide is predicted correctly as not at risk, this is a true negative.

D. **False Negative (FN):** The number of instances where the model incorrectly predicted the negative class (e.g., incorrectly predicting an individual is not at risk of suicide when they actually are at risk). It highlights how many actual positives were missed by the model.

   Example: If an individual who is at risk of suicide is predicted to be not at risk, this is a false negative.

Precision: Precision is the proportion of true positive predictions out of all positive predictions made by the model.

- **Pros:** Precision focuses on the accuracy of positive predictions, making it useful in scenarios where the cost of false positives is high.
- **Cons:** Precision can be misleading when the model rarely predicts the positive class, even if the model's recall is low.
- **Example:** In a suicide prediction model, if the model identifies 5 individuals as at risk (positive class) and only 2 of them are actually at risk, Precision = 2/5 = 40%. While this seems like reasonable precision, it means that 60% of the individuals identified as at risk were false positives, potentially leading to unnecessary interventions or stress for those individuals.
- **Takeaway:** Precision alone does not account for false negatives, which is critical in imbalanced datasets like suicide prediction, where the goal is to identify as many true positives as possible without overburdening the system with false alarms.

3) Recall: Recall (Sensitivity or True Positive Rate) is the proportion of actual positive cases that the model correctly identifies.

- **Pros:** Recall is crucial when missing positive cases (e.g., at-risk individuals) has severe consequences.
- **Cons:** High recall can come at the expense of precision, leading to many false positives.
- **Example:** Using the same suicide prediction model, if there are 10 actual at-risk individuals, and the model correctly identifies 8 of them, Recall = 8/10 = 80%. However, if this recall is achieved by labelling a large number of non-risk individuals as at-risk (lowering precision), the model might lead to unnecessary interventions and resources being used.
- **Takeaway:** In suicide prediction, a high recall is often desired, but it must be balanced with precision to avoid overwhelming systems with false positives.

4) F1 Score: The F1 Score is the harmonic mean of Precision and Recall, providing a balance between the two metrics. i.e. recall=(2*precision*recall)/(precision+recall)

- **Pros:** The F1 Score is particularly useful in imbalanced datasets, as it gives a single measure of a model's performance that considers both false positives and false negatives.
- **Cons:** The F1 Score might not fully capture the trade-offs between precision and recall in certain scenarios and can sometimes obscure the individual contributions of each but still better metric than previously discussed ones.

- **Example:** Suppose the precision of your suicide prediction model is 40%, and the recall is 80%. The F1 Score would be 2 * (0.4 * 0.8) / (0.4 + 0.8) = 53.3%. This F1 Score provides a more balanced view of the model's performance, considering both how many at-risk individuals are correctly identified and how many false alarms are raised.
- **Takeaway:** The F1 Score is a valuable metric for assessing model performance in imbalanced datasets like suicide prediction, especially when you need to balance the importance of precision and recall.

5) Area Under the Precision-Recall Curve (AUC-PR):AUC-PR measures the area under the curve of the Precision-Recall plot, which shows the trade-off between precision and recall across different thresholds.

- **Pros:** AUC-PR is particularly useful for evaluating model performance in imbalanced datasets, as it focuses on the performance of the model in identifying the positive class (e.g., at-risk individuals), which is often the minority class.
- **Cons:** AUC-PR can be less informative if the dataset is not highly imbalanced or if the precision and recall are both high across all thresholds.
- **Example:** In a suicide prediction model, if the AUC-PR is 0.75, this indicates that the model has a reasonably good ability to balance precision and recall across different thresholds. For instance, if the model has high precision at high recall levels, the AUC-PR value reflects how well the model performs in distinguishing at-risk individuals from non-risk individuals.
- **Takeaway:** AUC-PR provides a better assessment of model performance for imbalanced datasets compared to AUC-ROC, as it focuses specifically on the performance with respect to the positive class, making it highly relevant for tasks like suicide prediction.

Conclusion

The development of an AI-based suicide prediction model using binary classification marks a significant breakthrough in mental health intervention. This document has outlined the critical steps required to design, implement, and evaluate such a model, with a strong focus on building a robust dataset and upholding ethical standards. By integrating various data sources, the model aims to identify subtle patterns that signal suicidal risk, enhancing its ability to accurately classify individuals as either at risk or not. The complexity of training this model, due to the rarity of suicide cases, necessitates specific strategies to manage data imbalance and ensure reliable performance, with careful attention to minimizing false positives and negatives.

Ethical considerations are fundamental to this project, with principles such as informed consent, data anonymization, transparency, and cultural sensitivity playing a crucial role in maintaining the integrity of the data collection process and safeguarding participants' rights. The document emphasises the importance of ethical review and ongoing monitoring to ensure that the model development adheres to high privacy and respect standards. Ultimately, the AI suicide prediction model has the potential to transform mental health care by offering a proactive tool for identifying individuals at risk and enabling timely interventions.

Its success will depend on a balanced approach that combines technical expertise with ethical rigor, ensuring the model not only advances research but also serves as an effective and compassionate means of suicide prevention.