

Gustavo Adolfo Tabima Luque - 201914586  
Juan Carlos Eduardo Nunes Ariza-202010826  
César Luis Moreno González - 201924899

## Documento inicial de Entendimiento del proyecto:

Oportunidad / Problema de Negocio	La oportunidad es poder prevenir a través de predicción de calidad el suicidio. El problema de negocio es que algunos usuarios de reddit se están suicidando y no sé están detectando los casos de posible suicidio entre los usuarios de los Reddits con depresión a tiempo.
Enfoque analítico	Se necesita analizar los textos para determinar adecuadamente cuál es un relato de suicidio que permita determinar si el relato es en realidad un post donde se pide ayuda por parte de la víctima y separarlo de un post en broma (shitpost)
Organización y rol dentro de ella que se beneficia con la oportunidad definida	Se beneficia la comunidad de Reddit, en especial sus usuarios, que pueden recibir alertas y acompañamiento de parte de otras personas para ayudarlos en caso de necesitarlo, también ayuda a los moderadores del Thread sobre depresión para monitorear y tomar acciones ante sus usuarios
Técnicas y algoritmos a utilizar	Se tiene planeado un árbol de decisión que permita determinar que casos un post de reddit sea un reddit determinante para un suicidio. También se planea usar KNN y Support Vector Machine, Baggin, AbaBoost, Bernoulli Se van a procesar los datos tanto corrigiendo elementos de typos, asciis y números, como aplicando Stems y Lemmatización, para así poder generar tokens que puedan ser procesador por una vectorización y generar así un análisis completo del lenguaje.

## Definición de Roles para el proyecto:

- **Gustavo Tabima: Líder de Proyecto y Líder de Negocio**
- **César Luis Moreno Ronzales: Líder de Datos**
- **Juan Nuñez: Líder de Analítica**

**Tabla de horas y puntuaciones:**

<b>Nombre</b>	<b>Actividades</b>	<b>Horas</b>	<b>Puntuación</b>
<b>Gustavo Tabima</b>	<b>Generación Modelos:</b>  <b>Árboles de decisión</b>  <b>Random Forest</b>  <b>Entendimiento de datos</b>  <b>Desarrollo de entregables</b>  <b>Desarrollo de Wiki</b>	<b>20 Horas</b>	<b>33.3</b>
<b>César Moreno</b>	<b>Generación modelos:</b>  <b>Vector Machine</b>  <b>Bayessian Classifier</b>  <b>Bernoulli Classifier</b>  <b>Procesamiento de los datos</b>	<b>20 Horas</b>	<b>33.3</b>

	<b>Visualizacion de los resultados</b>		
<b>Juan Nunes</b>	<b>Generación de Modelos: KNN</b>  <b>Abbaboost</b>  <b>Presentación</b>	<b>20 Horas</b>	<b>33.3</b>