**Proyecto de Machine Learning**

**Problema:**

Dado un chat saber si en algún momento uno de los usuarios comienza a escribir diferente ya sea por una razón u otra.

Como sub-problema se tiene que dado un texto saber cuál de los integrantes del chat escribió dicho texto.

**Modelación del problema:**

Dado un chat se seleccionaron una cantidad de mensajes que comienzan en algún momento. En un diccionario se guarda por cada usuario todos sus mensajes. Haciendo uso de la biblioteca nltk de Python se realiza un procesamiento de texto donde por cada mensaje se va a eliminar las stopwords, o sea, las palabras que no aportan ningún significado, los signos de puntuación y se lematiza el texto, se llevan las palabras a su raíz gramatical. Luego de este procesamiento se vectorizará cada mensaje usando el TfidfVectorizer de la biblioteca sklearn y así se obtendrán los vectores de cada usuario por cada mensaje.

**Algoritmos de aprendizaje no supervisado usados en la resolución del problema:**

**- One Class SVM:**

Este estimador es sensible a valores atípicos, por lo tanto, no funciona muy para la detección de valores atípicos. One Class SVM es más adecuado para la detección de novedades cuando el conjunto de entrenamiento no está contaminado por valores atípicos. La detección de valores atípicos en alta dimensión es desafiante y One Class SVM podría dar buenos resultados en estas situaciones dependiendo del valor de sus hiperparámetros.

**- Robust Covariance**

**- Isolation Forest**