**Localización Por Potencias WI-FI**

El siguiente informe se encarga de explicar el proceso del modelo encargado de predecir la ubicación de un individuo u objeto dentro de un edificio de gran tamaño. Ya que generalmente los GPS no funcionan dentro de estos establecimientos, se busca la manera más eficiente para poder dar con la ubicación por medio de potencias de WI-FI, usando un set de datos que cuenta con los fingerprints de un centro comercial. Se generara múltiples modelos predictivos buscando dar con la ubicación más certera por medio de una comparación de los desempeños de cada modelo, se mostrara cual es el más óptimo con una breve explicación del mismo. Esto será concluido con recomendaciones para mejorarlo y mejorar el objetivo.

**Proceso**

El data set con el que se cuenta contiene un numero gigante de potencias de WI-FI, además coordenadas y otras puntos de referencias para dar con las ubicaciones dentro del edificio.  
La idea fue usar esas potencias para predecir las coordenadas como la longitud y latitud. También usar las mismas para dar con los otros puntos de referencias como el piso, si está o no dentro de un establecimiento y el tipo de área en el que se encuentra. Todos estos de manera individual.  
La Longitud y Latitud fueron predecidos por medio de un modelo de regresión.  
El ID del espacio, el piso y el edificio fueron predecidos por medio de un modelo de clasificación.  
Todos estos modelos se predijeron de manera individual solamente usando las potencias de WI-FI para cada uno, ya que dentro del establecimiento la aplicación solamente leerá las potencias para dar con la ubicación.  
Al set de datos se le convirtieron todos los valores de 100 que no representaban nada a 0.  
También se pasaron todos los valores negativos a positivos.  
Ya que el set de datos era muy grande, se utiliza solo un 10% del total de los datos para entrenarlos, esto con la idea de no hacer el modelo tan lento pero que si de un resultado certero.  
Las diferentes características a predecir debían estar en el tipo de valor correcto, ya fuera como un carácter o factor en los modelos de clasificación y los modelos de regresión el valor fue numérico.  
Se realizaron varios modelos como Random Forest, Support Vector Machine, C5.0 y entre otros para ver cuál daba con el mejor resultado.  
Estos modelos fueron comparados por medio de una matriz de confusión al igual que con la función resample.  
El modelo utilizado en todos los casos fue el Random Forest, ya que en las comparaciones da con el mejor resultado de Kappa y Accuracy. Además, con este mismo modelo se consiguieron los mejores RMSE y R2.

**Problemas con la Data**

El set de datos era gigante.  
El número de valores relativos para la predicción eran mínimos.  
Había varias características que no eran útiles para el modelo.

**Soluciones**

Se hizo una partición de datos inteligente, que no fuera muy pesado pero que diera un buen resultado.  
Solo se utilizaron los valores de importancia.  
Se eliminaron características innecesarias.

**Comparación de modelos**











Tomando en cuenta todos los números de los desempeños se muestra claramente porque el Random Forest es el modelo predictivo utilizado para generar la predicción. Este además es un código bastante adaptable y funcional para diferentes utilidades lo que lo hace muy versátil y certero.

Meta

Se logró generar múltiples modelos predictivos para cada factor que determina una ubicación.   
Cada uno de estos factores cuenta con una precisión decente, todos estos en conjunto dan un resultado muy confiable con respecto a la ubicación.  
Estos se pueden analizar por medio de matrices de confusión.

Datos  
  
Es posible que haya diferentes inconvenientes con datos, como que los puntos de acceso hayan dejado de funcionar y no se haya detectado la baja. Además, los datos recolectados pueden ser sustraídos con facilidad en caso de no haber la seguridad necesaria. Con el tiempo el modelo podría irse des actualizando y dejar de ser preciso. Por lo tanto.

Estos datos serán almacenados en servidores con gran capacidad y seguridad informática para su respectivo cuidado.  
Los datos serán recolectados por medio de cada celular y la aplicación que estén utilizando.  
Los puntos de acceso de wi-fi serán constantemente evaluados para saber que todos estén dando la señal óptima.   
Los algoritmos tendrán mantenimiento constante para mejorar el modelo, hacerlo más exacto y que este sea moldeado a las necesidades del momento.

Recomendaciones

Aunque el modelo cuenta con un nivel alto de precisión, y para que haya menos margen de error, cada factor que de una ubicación es determinado de manera individual para que estos en combinación puedan dar mayor confianza, siempre abre la oportunidad en la mejora, mas con la velocidad con la que se mueve el mundo, lo cambiante que es y las nuevas innovaciones que cada día hacen todo más sencillo.  
La sugerencia como analista de datos y experto en la materia, seria combinar la predicción con potencias con señales de Bluetooth, esta combinación daría un nivel altísimo de exactitud en la ubicación del objeto o individuo. Solo se deben instalar Faros de energía bluetooth en los diferentes las diferentes tiendas del comercio. La aplicación daría la ubicación dependiendo la cercanía del faro, esto junto al detector de potencias daría el máximo desempeño posible.  
Existe otro método muy común llamado localización y mapeo simultáneo, este consiste en grabar todo el área por medio de drones y robots con cámaras, abarcar todo el espacio y por medio de cámaras instaladas en diferentes ángulos, determinar la ubicación. Ya que esto daría un giro al proyecto la recomendación seria mejorar por medio de señales bluetooth.