Jalisco Connection Points

Proyecto Programación para Análisis de Datos

Esteban Javier Berumen Nieto

Iteso  (Instituto Tecnológico y de Estudios Superiores de Occidente)

Introducción

En este proyecto, se emplea el algoritmo KNN para predecir el ancho de banda contratado en 2014, por un listado de instituciones públicas de Jalisco en el año 2016. El objetivo principal es evaluar la precisión del modelo en función de las características de estas instituciones.

Los datos fueron obtenidos del sitio <https://datos.jalisco.gob.mx/dataset/puntos-de-conexion-ejalisco> en donde encontramos un dataset con las siguientes variables incluidas: **[clave de inmueble, tecnologia instalada, ancho de banda contratado 2014, institucion, nombre del centro, turno/horario, nivel, region, municipio, localidad, domicilio, codigo postal, longitud, latitud]** en donde encontramos 6716 registros.

Desarrollo

Dentro del preprocesamiento de datos se realizó una limpieza en donde todos los datos faltantes y N/D del dataset fueron remplazados por la primera moda de la columna correspondiente, esto ya que todas las columnas del dataset son categóricas, esto a su vez hiso que fuera necesario usar el *OrdinalEncoder* de la librería *sklearn*, para poder realizar la codificación de las variables, además se dividió el data set e un train y test en donde el train es el 20% del original y test es el 80%.

Gráfico

Descripción generada automáticamenteGráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente A continuación, se eliminaron algunas columnas debido de diferentes problemas u objetivos; como lo son las columnas de: **clave de inmueble, ancho de banda contratado 2014, nombre del centro, longitud, latitud**. En el caso del ancho de banda, esta columna se convierte en el target. También se realizaron los histogramas de las diferentes columnas, así como un mapa de correlación de las variables.

Gráfico

Descripción generada automáticamenteGráfico, Gráfico de líneas

Descripción generada automáticamente También se realizaron histogramas para el traget tanto en la parte del train como en la parte del test.

Gráfico, Gráfico de rectángulos

Descripción generada automáticamenteGráfico, Gráfico de rectángulos

Descripción generada automáticamente Mapa de correlación de variables.