Equipo Kaggle: **Sinapsis**

Integrantes:

**Alejandra Lelo de Larrea Ibarra 000124433**

**Luis Fernando Cantú Díaz de León 000125755**

Proyecto Final

**Sistema para compartir bicicletas: pronóstico según las condiciones climáticas y ambientales.**

Los sistemas para compartir bicicletas (como EcoBici en México) son una nueva generación del tradicional alquiler de bicicletas. En éstos todo el proceso de membresía, alquiler y devolución se ha automatizado. Hoy en día, existe un gran interés en estos sistemas debido a su importante papel en la reducción del tráfico vehicular, el cuidado del medio ambiente y la reducción en los problemas de salud. Actualmente, hay alrededor de 500 programas para compartir bicicletas en todo el mundo que se componen de más de 500 mil bicicletas.

A través de estos sistemas, el usuario puede alquilar una bicicleta fácilmente desde una estación particular y regresarla en una estación que no tiene por qué coincidir con la inicial. Opuesto a otros servicios de transporte como el autobús o el metro, la duración del viaje, la posición de salida y la de llegada se registran explícitamente en estos sistemas. Esta característica convierte el sistema de intercambio de bicicletas en una red de sensores virtuales que se puede usar para detectar la movilidad en la ciudad y otras características. Por ejemplo, el proceso de alquiler de bicicletas compartidas está altamente correlacionado con las condiciones ambientales. Las condiciones climáticas, la precipitación, el día de la semana, la estación del año, la hora del día, etc. pueden afectar las conductas de alquiler.

De esta manera, se busca predecir el número de bicicletas rentadas por hora según las condiciones climáticas y ambientales y comparar los resultados obtenidos con distintos métodos vistos en clase. Para ello, se utiliza el registro histórico (años 2011 y 2012) del sistema Capital Bikeshare, Washington D.C., EE. UU.[[1]](#footnote-1) La base de datos contiene 17,379 observaciones para 17 variables; entre ellas se encuentran: la fecha y hora, estación del año, si el día es laborable o no, tipo de clima, condiciones ambientales (temperatura, humedad, etc) y número de bicicletas rentadas (variable de interés). Para separar las muestras, se tomarán los datos por hora del 1° de enero del 2011 al 31 de diciembre del 2011 (aprox. 50% de las observaciones) para entrenamiento, del 1° de enero del 2012 al 30 de junio del 2012 (aprox. 25% de las observaciones) para validación y del 1 de julio del 2012 al 31 de diciembre del 2012 (aprox. 25% de las observaciones) para prueba. Los métodos a utilizar para la estimación y comparación son: regresión lineal, regresión lineal con regularización, redes neuronales, bosques aleatorios y gradient boosting machine.

1. Los datos fueron recuperados de UCI Machine Learning Repository <https://archive.ics.uci.edu/ml/datasets/bike+sharing+dataset> [↑](#footnote-ref-1)