



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

**MODELO PREDICTIVO DE CONSUMO ENERGÉTICO EN AUTOMÓVILES
ELÉCTRICOS EN BASE A MÉTODOS DE APRENDIZAJE DE MÁQUINA**

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA INGENIERÍA,
MENCION ELÉCTRICA

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO

CÉSAR IGNACIO BAEZA FERNANDEZ

PROFESOR GUÍA:

Marcos Orchard Concha

PROFESOR CO-GUÍA:

Claudio Burgos Mellado

MIEMBROS DE LA COMISIÓN:

Jorge Silva Sánchez

Williams Calderón Muñoz

SANTIAGO DE CHILE

2023

e

RESUMEN DE LA TESIS PARA OPTAR AL GRADO DE MAGÍSTER
EN CIENCIAS DE LA INGENIERÍA, MENCIÓN ELÉCTRICA

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO

POR: CÉSAR IGNACIO BAEZA FERNANDEZ

FECHA: 2023

PROF. GUÍA: Marcos Orchard Concha

PROF. CO-GUÍA: Claudio Burgos Mellado

MODELO PREDICTIVO DE CONSUMO ENERGÉTICO EN AUTOMÓVILES ELÉCTRICOS EN BASE A MÉTODOS DE APRENDIZAJE DE MÁQUINA

Debido al cambio climático y la urgencia de reducir las emisiones de carbono, los vehículos eléctricos (EV) son una de las alternativas más populares para revertir esta situación. Con el fin de aumentar la adquisición de estos EV y los problemas relacionados a *range anxiety*, el objetivo de la tesis es estimar de energía dado una ruta, utilizando información histórica. Para esto se realizó un estudio fenomenológico de todas las variables que se ven involucradas en el consumo y se utilizó herramientas basadas en aprendizaje de máquina para lograr esta estimación. El modelo consta de 3 grandes etapas, la primera un modelo basado en una red recurrente *Long Short Term Memory*, que a diferencia de las redes convencionales, permite procesar secuencias de datos, para en este caso, estimar la distribución de velocidad, el segundo modelo es la estimación de consumo por tramo utilizando un árbol de regresión y finalmente el modelo que integra la salida de la red en la entrada del árbol de regresión, obteniendo una distribución de estimación de consumo por tramo, el cual se integra para obtener la distribución de consumo de toda la ruta, obteniendo una mejora de cerca de un 85 % en la estimación sobre artículos recientes.

*Nadie sabe el resultado,
así que lo único que nos queda es tomar una decisión de la cual no nos arrepintamos.*

Levi Ackerman

Agradecimientos

No tengo más que agradecer a cada una de las personas que fueron parte de este camino. A mis padres, César y Soledad, por darme todo lo que han podido y más, por apoyarme en los momentos más difíciles y celebrar en cada pequeño logro. A mis hermanos Fabián, Adolfo y Damian, por darme su cariño y compartir risas siempre que se podía. Gracias a todos ustedes, los que conforman esta familia, buena para discutir y molestar, pero siempre estar allí cuando uno necesita un consejo o simplemente conversar (o copuchar). Agradecer a mi abuela Zina por ser parte de mi vida y estar presente desde que tengo memoria. A mi abuelo Reinaldo, por las risas y conversaciones los fin de semana.

Quiero agradecer a los bachis, al grupo que se conformó con estudio y carrete en Juan Gomez Millas, por hacer la vida universitaria más interesante, con las cervecitas y los perreos hasta el el -1. También a los amigos que formé en beacuhef y siguen estando presentes. A los que conocí en eléctrica y los que el G.O. de escalada donde realizamos iniciativas muy lindas en torno a un deporte que nos apasiona.

También quiero agradecer a los "The Rial", amigos muy importantes para mi, los que siempre íbamos a las juntas, de los cuales tengo una bonita amistad.

Quiero dedicar unas palabras especiales a Darinka, una amiga muy especial, en quien confié mucho y me han escuchado en momentos difíciles. Además agradecer a Pancho y Pipe, muy buenos amigos, con quienes he compartido pegues muy buenos, aventuras y borracheras, gracias por el apoyo y la buena onda. Los quiero.

Dedicar un párrafo especial a Naroa, mi pareja, la persona que más me ha apoyado la mayor parte de este proceso, una persona maravillosa, gracias por todas las aventuras y por estar allí cada vez que necesito tu apoyo o consejo, gracias por ser tan apañadora y por todo lo que hemos construido juntos. Ojala todos tengan una Naroa en su vida. Sigamos creciendo juntos, te amo.

También agradecer a profesor Marcos Orchard por ser un gran académico, que me acompañó en este camino, además de ser una gran persona. A Claudio Burgos por apoyarme en el proceso de la tesis como profesor co-guía y al comité de revisión por el tiempo dedicado.

Agradecer a los que están y los estuvieron, a los de siempre y los de aveces, sin ustedes esto no habría sido lo mismo.

Tabla de Contenido

1. Introducción	1
1.1. Motivación	1
1.2. Hipótesis	3
1.3. Objetivos Generales	3
1.4. Objetivos Específicos	3
1.5. Alcance	3
2. Marco Teórico	4
2.1. Antecedentes de Vehículos Eléctricos	4
2.2. árboles de Regresión	4
2.3. Partición Binaria Recursiva	5
2.4. Podado	5
2.5. Boosting	6
2.6. GradientBoosting	6
2.7. XGBoost	7
2.8. Light Gradient Boosting Machine (LightGBM)	8
2.9. DeepLearning	11
2.10. Redes Recurrentes	11
2.11. Long Short-Term Memory Recurrent Neural Network (LSTM)	13
2.12. Incertidumbre en Red LSTM	14
2.13. <i>SHapley additive exPlanations</i> (SHAP)	14
3. Estado del Arte	16
3.1. Determinación de los principales factores que influyen en el consumo de vehículos eléctricos	17
3.2. Fenomenología del Consumo en Vehículos Eléctricos	18
3.2.1. Métodos de estimación de aceleración	20
3.2.1.1. Evaluación	21
3.2.2. Soluciones basadas en Machine Learnig	24
3.2.2.1. <i>Energy consumption estimation in electric vehicles considering driving style</i>	24
3.2.2.2. <i>A Data-Driven Method for Energy Consumption Prediction and Energy-Efficient Routing of Electric Vehicles in Real-World Conditions</i>	24
3.2.2.3. <i>Energy consumption analysis and prediction of electric vehicles based on real-world driving data</i>	25
4. Metodología	28

4.1. Caso de Estudio	28
4.2. Base de Datos	28
4.3. Extracción de Características	30
4.3.1. Segmentación de Ruta	32
4.4. Modelo de Estimación de Velocidad	32
4.4.1. Estructura de características de modelo	34
4.4.2. Creación de Modelo	35
4.5. Modelo de Consumo de Energía	36
4.5.1. Estructura de Características del Modelo	36
4.5.2. Creación de Modelo	36
4.6. Modelo Integrado	36
5. Resultados y Discusiones	39
5.1. Modelo de Velocidad	39
5.2. Modelo de Consumo	46
5.2.1. Análisis de Integral de Consumo	50
5.3. Modelo Integrado	52
5.3.1. Modelo de Sobre-Estimación de Consumo	58
6. Conclusiones y Recomendaciones	62
6.1. Trabajos Futuros	63
Bibliografía	64
Anexos	67
Anexo A. Ficha Técnica EV Nissan Leaf 2013	67
Anexo B. Estructura DataFrame	68
Anexo C. Modelo de Red LSTM	69
Anexo D. Análisis Proceso de MonteCarlo	70
Anexo E. Correlación Entre Periféricos y Condiciones Ambientales	72

Índice de Tablas

3.1.	Caracterización de los parámetros de requerimiento de potencia. Adaptado de (Kim, Lee, y Shin, 2013)	20
3.2.	Resultados de metodología implementada, adaptado de (Kim et al., 2013) . . .	23
3.3.	Resultados de metodología utilizada, adaptado de (Felipe, Amarillo, Naranjo, Serradilla, y Díaz, 2015)	24
3.4.	Resultados obtenidos, adaptado de (Cauwer, Verbeke, Coosemans, Faid, y Mierlo, 2017)	25
4.1.	Tabla de análisis de rango utilizando modelo de consumo	33
4.2.	Error de red LSTM utilizando distintos regresores	35
5.1.	Análisis de error por tramos de velocidad en conjunto de entrenamiento	42
5.2.	Análisis de error por tramos de consumo	49
A.1.	Especificaciones técnicas Nissan Leaf 2013	67
B.1.	Estructura de dataframe, el cual es normalizado para el posterior entrenamiento del modelo de estimación de velocidad.	68
C.1.	Modelo de red LSTM utilizado para el proceso de estimación de Velocida . . .	69
D.1.	Análisis de nro de iteraciones en proceso de Monte Carlo	70

Índice de Ilustraciones

2.1.	Diagrama de modelo de Velocidad (Krzywinski y Altman, 2017)	5
2.2.	Diagrama de métodos de aprendizaje <i>Level-wise</i> y <i>Leaf-wise</i> (Fan et al., 2019).	8
2.3.	(A)Histogramas (B) Gradient-based one-side sampling (C) EFM. Diagramas de algoritmos utilizados en LightGBM para acelerar el entrenamiento.(Gan et al., 2021).	9
2.4.	Diagrama de Red Neuronal con 2 capas ocultas y múltiples salidas, utilizado en el problema de <i>MNIST handwritten-digit</i> , la entrada tiene 783 unidades, las siguientes dos capas tienen $K_1 = 256$ y $K_2 = 128$ unidades respectivamente(Daniela, Trevor, Gareth, y Witten, 2021).	12
2.5.	Diagrama de una RNN simple, donde la secuencia de entrada es $\{X_l\}_1^L$ y la salida es única(Daniela et al., 2021).	12
2.6.	Diagrama de una celda LSTM (Chevalier, 2018).	13
2.7.	Ejemplo de red U-Net 3D con aplicación de <i>dropout</i> en las capas de decodificación, tanto en el entrenamiento como en la validación. (Avci et al., 2021) . . .	15
3.1.	Clasificación del estado del este de estimación de consumo en EV, adaptado de (Y. Chen et al., 2020).	17
3.2.	Diagrama Ishikawa de los factores que impactan en el consumo eléctrico de un EV, adaptado de (Li, Stanula, Egede, Kara, y Herrmann, 2016).	19
3.3.	Fuerzas involucradas en el requerimientos de potencia (Kim et al., 2013). . . .	19
3.4.	Estructura de sistema de predicción de requerimiento de potencia (PRP), extraído de (Kim et al., 2013)	22
3.5.	La ruta de autopista US-101, Los Angeles, USA, en dirección sur. Extraído de (Kim et al., 2013)	22
3.6.	Error inherente por estimación discreta, extraído de (Kim et al., 2013)	23
3.7.	Diagrama de solución propuesta, adaptado de (Cauwer et al., 2017)	25
3.8.	Diagrama de framework basado en ML para estimación de energía(Zhang, Wang, Liu, y Zhang, 2020)	26
3.9.	Resultados del RMSE y MAE para la metodología propuesta y la convencional (Zhang et al., 2020)	27
4.1.	Mapa de elevación de la zona de Ann Arbor, Michigan, USA.	29
4.2.	Precipitaciones año 2018 en Ann Arbor, Michigan, USA (Weatherspark, 2022).)	29
4.3.	Temperaturas año 2018 en Ann Arbor, Michigan, USA (Weatherspark, 2022))	30
4.4.	El gráfico superior muestra la distribución del largo de los viajes y la figura inferior la distribución de el total recorrido por mes(Oh, Leblanc, y Peng, 2019)	31
4.5.	Diagrama de extracción de características general	32
4.6.	Ejemplo de segmentación de ruta en un rango $\in [50$ y $150[$ metros.	33
4.7.	Ruta de ejemplo	34

4.8.	Diagrama del proceso de entrenamiento del modelo de velocidad utilizando una red LSTM. El parámetro remarcado indica el target que se quiere estimar. . .	35
4.9.	Diagrama del proceso de entrenamiento del modelo de consumo utilizando un árbol de regresión <i>LightGBM</i> . El parámetro remarcado indica la variable a estimar	37
4.10.	Diagrama del proceso de validación donde se unen el modelo de estimación de velocidad y el de consumo de energía.	38
5.1.	Error de modelo de velocidad sobre todo el conjunto de validación.	40
5.2.	Distribución de error según rango	41
5.3.	Distribución de error según rango	43
5.4.	<i>SHAP values</i> de todas las variables sobre el modelo de velocidad.	44
5.5.	<i>SHAP values</i> de variables asociadas al clima sobre modelo de velocidad.	45
5.6.	<i>SHAP values</i> de variables artificiales de la ruta.	46
5.7.	Distribución de error según rango	47
5.8.	Error de modelo de consumo sobre todo el conjunto de validación (caso base), utilizando variable de velocidad real.	48
5.9.	(a) Distribución de error para freno regenerativo debido a los efectos del freno regenerativo (b) Consumo negativo debido a los cambios de velocidad positivos	48
5.10.	<i>SHAP values</i> modelo de consumo de energía	50
5.11.	Integral del consumo, target vs estimación, para el modelo base con velocidad entregada por data	51
5.12.	Error de modelo de consumo acumulado sobre todo el conjunto de validación (caso base), utilizando variable de velocidad real.	52
5.13.	Distribución de modelo consumo instantáneo punto a punto, utilizando velocidad estimada a la salida dl modelo de velocidad.	53
5.14.	Media de las distribución para el interval de confianza y el valor esperado según la distribuciones de la figura ??	55
5.15.	Error de modelo de consumo sobre todo el conjunto de validación, utilizando variable de velocidad estimada.	56
5.16.	Error de estimación de consumo utilizando la velocidad estimada por el modelo anterior. (a) Distribución de error para freno regenerativo debido a los efectos del freno regenerativo (b) Consumo negativo debido a los cambios de velocidad positivos	56
5.17.	Integral del consumo esperado para la distribución obtenida en la figura ??, obteniendo la distribución del consumo total.	57
5.18.	Error de modelo de consumo acumulado sobre todo el conjunto de validación, utilizando variable de velocidad rea y modelo entrenado con q0.5 (mediana)	58
5.19.	Resultados de en viajes seleccionados del conjunto de test, utilizando un modelo entrenado para estimar el quantile 0.4	59
5.20.	Error de modelo de consumo acumulado sobre todo el conjunto de validación, utilizando variable de velocidad real y modelo entrenado utilizando q0.4	60
D.1.	CDF de la distribución de velocidades utilizando distintos numero de procesos de Monte Carlo	71
E.1.	Correlación entre consumo de aire acondicionado y calefacción y variable humedad y temperatura.	72

Capítulo 1

Introducción

1.1. Motivación

Debido al cambio climático y la urgencia de reducir las emisiones de carbono, los vehículos eléctricos (EV) son una de las alternativas más populares para revertir esta situación. La literatura informa que las ventas en el mercado global de este tipo de vehículos han aumentado de 2,26 millones de unidades en 2019 a un total de 3,24 millones durante 2020 (Irle, 2020). Además, cuando los datos de las ventas de 2010 a 2018 se utilizan para definir un modelo de crecimiento logístico, los resultados del pronóstico muestran que para el año 2032 todos los vehículos de pasajeros representarán cerca del 30 % de la flota (Rietmann, Hügler, y Lieven, 2020).

Por otro lado, la introducción de vehículos eléctricos ha sido notoria en los últimos cinco años en América Latina. Por ejemplo, entre 2018 y 2019, Chile tuvo un aumento del 68 % en la cantidad de vehículos eléctricos, y actualmente, hay una flota aproximada de 900 unidades. Otros países, como Panamá, han tenido un proceso más lento, pero aún así, se puede evidenciar la presencia de vehículos eléctricos (ONU Environment Programme, 2019) (Javier Ochoa, s.f.).

Los desarrollos recientes en la tecnología de almacenamiento de iones de litio (Li-ion) han desencadenado una nueva revolución en la industria automotriz, aunque los vehículos eléctricos (EV) se ven afectados por la combinación más diversa de condiciones ambientales y de conducción que afectan el rango de autonomía y el rendimiento general. Dos de los indicadores más utilizados cuando se utiliza cualquier tipo de batería son el estado de carga (SOC) y el estado de salud (SOH). Sin embargo, la estimación de SOC no es muy útil para pronosticar la autonomía, ya que esta variable depende no solo de SOC si no también de características intrínsecas de la ruta a seguir así también como el tipo de conducción (Cauwer et al., 2017). Esto se puede evidenciar en el estudio de *DarioPevce*, donde queda en evidencia que bajo de un 30 % de SOC las personas se sienten poca seguridad de poder llegar a su destino, aunque el sistema marque 50km de autonomía, ya que no se tiene certeza de la autonomía real, este fenómeno se conoce como *Range anxiety* (Pevce et al., 2019).

A pesar de que la estimación de SOC se utiliza como estimador de autonomía en la mayoría de los casos, este no da información precisa, ya que la autonomía no solo se basa en el estado actual de batería si no también del estado de factores exógenos, como el tipo de manejo ,

inclinación, temperatura externa, humedad del camino, entre otros (Kim et al., 2013). Por otro lado un estudio realizado por investigadores de la Universidad de Chile, muestra que la cantidad de energía que puede entregar una batería no es fija y depende del perfil de uso (Rozas, Troncoso-Kurtovic, Ley, y Orchard, 2021). Es por esto, que este trabajo busca estimar la energía consumida dado una ruta, ya que sus aplicaciones tendrán un mayor impacto en toma de decisión futuras.

Con el fin de solucionar el paradigma de autonomía, es decir, si el estado actual de el pack de baterías es capaz de mover el vehículo al destino deseado, es que se han realizado diversos esfuerzos de estimar el consumo energético requerido por un EV tomando parámetros intrínsecos de la ruta a seguir. Entre los trabajos que se han realizado existen dos grandes metodologías, la primera basada en la fenomenología donde se busca modelar el comportamiento de los diversos sistemas involucrados en el movimiento de un EV utilizando modelación física y/o lógica difusa. Por otro lado gracias a la mejora de sensores es que se nacen modelos basados en en data, donde por lo general de trabaja con modelos tipo caja negra que se ajustan para lograr generalizar el consumo de un EV dado variables de entrada (Y. Chen et al., 2020).

Dentro de los modelos basados en fenomenología, se puede ser como existen diversos documentos que buscan ajustar los parámetros del consumo según los requerimientos físicos (Kim et al., 2013), este tipo de metodologías son poco escalables pero tienen la ventaja de que permiten entender el problema. Los años siguientes y con los avances en aprendizaje de máquina, se han planteado diversas soluciones basados en algoritmos de caja negra aplicando redes neuronales con el fin obtener perfiles de velocidad dado una ruta y regresores lineales para obtener una estimación del consumo total de la ruta (Cauwer et al., 2017), sin embargo la cantidad de datos que se tienen para entrenar los modelos no son lo suficientemente grandes para entrar modelos más complejos que tienen mejor desempeño.

En estos últimos años surgió un paper que tiene como objetivo generar una base datos a gran escala que toma la información de conducción de 383 vehículos conducidos en Ann Arbor, Michigan, USA (Oh et al., 2019). Los autores señalan que los datos son obtenidos con una alta tasa de muestreo y una obtención de latitud de longitud de alta resolución, por lo que por primera vez se tiene una base datos con datos de alta fiabilidad. Esto abre la puerta a poder tener una mejor solución al problema de estimación de consumo energético, utilizando información de alta precisión y gran escala que permite aplicar técnicas de aprendizaje de máquina más recientes donde se propone utilizar redes LSTM para la estimación de velocidad y árboles de regresión que estiman el consumo eléctrico tomando como variables de entrada variables asociadas a la fenomenología del sistema.

1.2. Hipótesis

Es posible modelar el consumo energético de una ruta arbitraria a través de un algoritmo de aprendizaje de máquina entrenado como datos asociados a segmentos de ruta con propiedades y características específicas, utilizando múltiples herramientas de aprendizaje de máquina, obteniendo la distribución de consumo para un EV en ruta y mejorando el desempeño de la estimación en comparación a los artículos actuales.

1.3. Objetivos Generales

Se busca generar un modelo basado en aprendizaje de máquina que logre modelar el consumo energético de un vehículo eléctrico, utilizando data histórica de 3 vehículos eléctricos que recorren Ann Arbor, Michigan, USA, durante 1 año.

1.4. Objetivos Específicos

- Identificar variables que ayuden a la estimación de consumo energético, basado en la fenomenología.
- Identificar variables que ayuden a la estimación de velocidad, basado en la fenomenología.
- Entrenar modelo de consumo energético basado en aprendizaje de máquina.
- Entrenar un modelo de estimación de velocidad basado en aprendizaje de máquina que mapear las variables de entrada a una distribución de estimación de velocidad.
- Generar modelo integrado de estimación de consumo utilizando las distribuciones obtenidas en el modelo de velocidad.

1.5. Alcance

El problema se divide en dos grandes aristas, la primera consiste en la modelación de consumo de energía por kilómetro cada 100 metros el cual se basa en el modelo físico de requerimiento energético, utilizando la velocidad y cambio de velocidad como principales características (Kim et al., 2013). El segundo problema consiste en lograr estimar la distribución de velocidad promedio cada 100 metros, dado la ruta. De esta forma se tiene un problema cerrado donde dado una ruta específica, se toman puntos geográficos para caracterizar la ruta y estimar la distribución de posibles velocidades, de esta forma se puede obtener una distribución de consumo utilizando el modelo previamente entrenado, para después realizar un estudio de los diferentes consumos en la ruta y así también la energía total requerida para completar la ruta.

Capítulo 2

Marco Teórico

2.1. Antecedentes de Vehículos Eléctricos

En los últimos años, el uso de vehículos eléctricos ha aumentado significativamente debido al esfuerzo de varias políticas gubernamentales para minimizar las emisiones de dióxido de carbono. Por ejemplo, una de las acciones es la introducción de leyes y regulaciones para el uso de vehículos eléctricos que brindan una variedad de beneficios a los usuarios, como descuentos de impuestos y estacionamiento público preferido (Gómez-Gélvez, Mojica, Kaul, y Isla, 2016). Como resultado, el mercado de vehículos eléctricos ha aumentado la tasa de crecimiento anual compuesta en un 26,8% entre los años 2021 y 2030 (Markets y Markets, 2021). Los lineamientos establecidos por varios países evidencian esta situación; tal es el caso de Chile, donde se espera que para el 2050 todo el transporte público sea cero emisiones (Gómez-Gélvez et al., 2016).

Los vehículos eléctricos se pueden clasificar en tres categorías según la naturaleza de la fuente de energía. Los tipos principales son: vehículos eléctricos de batería (BEV), vehículos eléctricos híbridos (HEV) y vehículos eléctricos híbridos enchufables (PHEV) (Gómez-Gélvez et al., 2016).

- BEVs: El motor se alimenta solo con una batería eléctrica.
- HEV: este tipo de automóviles tienen una batería eléctrica y un motor de combustión. La batería es cargada por el motor de combustión y ruptura regenerativa.
- PHEV: similar al HEV, pero en este caso la batería se puede cargar cuando se conecta a un cargador.

2.2. árboles de Regresión

Corresponde a una metodología de aprendizaje supervisado, que dado un set de datos con n ejemplos y m características. Donde cada árbol q busca dividir el espacio mediante nodos lógicos jerárquicos, donde se pregunta si $x_i < \alpha$ para cada observación x_i (Daniela et al., 2021). De esta forma se divide el espacio de decisión en umbrales excluyentes, ya que existe un único camino para llegar a cada hoja, como se muestra en la figura 2.1.

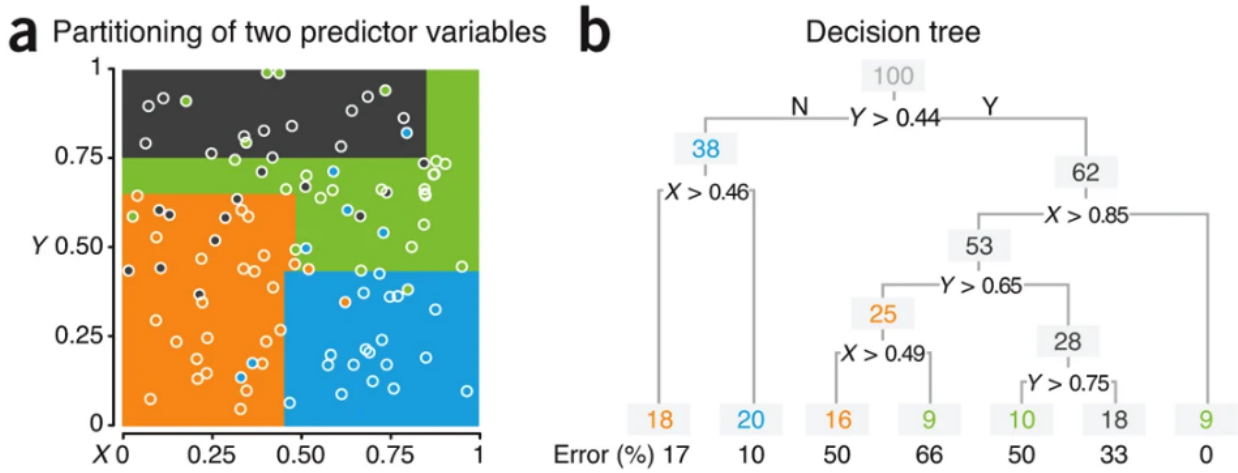


Figura 2.1: Diagrama de modelo de Velocidad (Krzywinski y Altman, 2017)

2.3. Partición Binaria Recursiva

Debido a que el esfuerzo computacional que se necesita para que cada árbol divida el espacio de decisión de forma óptima, resulta muy costoso, el algoritmo que se implementa es conocido como Partición Binaria Recursiva, donde se busca se busca minimizar en cada iteración se busca minimizar la suma de los errores al cuadrados (RSS por sus siglas en ingles)(Daniela et al., 2021) , cuya ecuación está dada por:

$$\sum_{j=1}^J \sum_{i \in P_j} (y_i - \hat{y}_{P_j})^2 \quad (2.1)$$

Donde \hat{y}_{P_j} corresponde la perdición para la partición j-esima del árbol q.

2.4. Podado

El proceso descrito anteriormente, se puede realizar de forma indefinida obteniendo buen resultado en el conjunto de entrenamiento, sin embargo, esto lleva a que el modelo se sobreajuste a la data, por lo que existen diversas 2 principales formas de solucionar esto:

- Pre-Podado: En este método se define una cantidad de hojas mínima β para realizar la partición de un subárbol, de esta forma se detiene la partición si no se cumple este criterio.
- Podado: A diferencia del *Pre-Podado*, el podado se realiza una vez que el algoritmo encuentra la mejor separación de espacio T_0 , donde el RSS es menor. Luego se busca los subárboles $T \subset T_0$ que minimizan la siguiente ecuación

$$\sum_{m=1}^{|T|} (y_i - \hat{y}_{R_m})^2 + \alpha |T| \quad (2.2)$$

conocida como *weakest link pruning*, donde α corresponde a un parámetro que se ajusta con el fin de encontrar la mejor relación entre desempeño y sobreajuste, cuando $\alpha = 0$ entonces $T = T_0$ (Daniela et al., 2021).

2.5. Boosting

Con el fin de mejorar el desempeño los árboles de regresión es que se implementa una metodología llamada *Boosting*, la cual consiste en entrenar árboles de forma secuencial, en donde el árbol q_i se encarga de ajustar los datos residuales de el árbol q_{i-1} con $i > 1$. Este método es más lento que el entrenamiento con un solo árbol, sin embargo, se obtienen mejores resultados con menor overfitting (Daniela et al., 2021).

2.6. Gradient Boosting

El algoritmo de Gradient Boost consiste en obtener una función $F(x) = y$ que defina una relación entre las variables x e y de un conjunto de datos $\{x_i, y_i\}_1^N$, que mapeando el espacio x a y , mejorando la función objetivo $F(x)$ que va mejorando al ir entrenando iterativamente, minimizando la función de perdida diferenciable $L(y, F(x))$ (Friedman, 2002). Los parámetros de $F(x)$ que se ajustan utilizando el entrenamiento utilizando data, donde se deben seguir los siguientes pasos:

1. Se inicializa el modelo con valores constantes: $F_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=0}^n L(y_i, F(x))$ donde γ a los valores predichos por el algoritmo.
2. Una vez inicializado $F(x)$, se itera sobre la cantidad de árboles definidos M .
 - Se calculan los pseudo-residuos r_{im} a partir del cálculo del gradiente.

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \quad (2.3)$$

para $i = 1$ hasta n

- Se ajusta el árbol de regresión con los valores r_{im} anteriormente calculado, luego se crean la regiones terminales R_{jm} para $j = 1 \dots J_m$ que corresponde a la cantidad de hojas.
- Se determinan las predicciones para J_m hojas de cada árbol m , donde se minimiza:

$$\gamma_{jm} = \underset{\gamma}{\operatorname{argmin}} \sum_{x_j \in R_{ij}} L(y_i, F(x)_{m-1}(x_i) + \gamma) \quad (2.4)$$

Donde $F_{m-1}(x_i)$ corresponde a la predicción del árbol $m-1$ para x_i .

- Utilizando los valores de γ que minimizan la función de costo, se obtiene la predicción del árbol m para cada entrada x_i de la siguiente forma:

$$F_m(x) = F_{m-1}(x) + \nu \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm}) \quad (2.5)$$

Donde ν corresponde a la tasa de aprendizaje que se utiliza para escalar los árboles y evitar sobreajuste, el cual es un valor en 0 y 1.

2.7. XGBoost

XGBoost corresponde a un algoritmo que se basa en el concepto de *Boosting*, donde dado un set de datos con n ejemplos y m características $D = (x_i, y_i) (|D| = n, x_i \in \mathbb{R}^m, y_i \in \mathbb{R})$ (T. Chen y Guestrin, 2016) correspondiente a un modelo de árbol como se muestra en la figura 2.1.b donde para representar una salida se realizan K sumas de funciones.

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (2.6)$$

Donde $F = f(x) = w_{q(x)}(q : \mathbb{R}^m \rightarrow T, w \in \mathbb{R}^T)$ es el espacio de árboles de regresión (*CART* por sus siglas en ingles), donde q corresponde a la estructura de cada árbol. T el numero de de hojas en el árbol y f_k corresponde a la estructura de cada árbol q con peso w . Para encontrar las funciones f_k , se minimiza la siguiente función objetivo:

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (2.7)$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$$

Donde l corresponde a la *differentiable convex loss function* que mide la diferencia entre \hat{y}_i y el valor real y_i , el termino Ω corresponde al termino de penalización por complejidad de árbol, para evitar overfitting (T. Chen y Guestrin, 2016). Debido a la complejidad de la ecuación 2.7, es que se necesita agregar el parámetro $f_t(x_i)$ a la función de costo.

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{t-1} + f_t(x_i)) + \Omega(f_t) \quad (2.8)$$

Por lo que escribiendo su aproximación de de segundo orden y eliminando los términos constantes, se obtiene.

$$\tilde{L}^{(t)} = \sum_{i=1}^n [g_i, f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) \quad (2.9)$$

Donde $g_i = \partial_{\hat{y}} l(y_i, \hat{y}_i^{(t-1)})$ y $h_i = \partial_{\hat{y}}^2 l(y_i, \hat{y}_i^{(t-1)})$. Por lo que optimizando la expresión se obtiene la siguiente representación para los pesos de cada hoja j w_j^* (T. Chen y Guestrin, 2016).

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (2.10)$$

Donde $I_j = \{i \mid q(x_i) = j\}$. Por lo que la función óptima toma el valor de:

$$\tilde{L}^{(t)} = -\frac{1}{2} \left(\frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} \right) + \gamma T \quad (2.11)$$

Esta expresión permite saber si dividir como afecta al desempeño general el dividir un subárbol en rama izquierda y derecha. Por lo que luego se haber generado K árboles se realiza la predicción utilizando la ecuación 2.6.

2.8. Light Gradient Boosting Machine (LightGBM)

LightGBM corresponde a un algoritmo de aprendizaje de máquina basado en el gradiente en el árbol de decisión utilizando *boosting* (Ke et al., 2017). Donde la mayor diferencia con XGBoost, es que LightGBM utiliza un algoritmo basado en histogramas para el entrenamiento lo que reduce considerablemente el tiempo de entrenamiento y reduce la memoria requerida, además de utilizar el algoritmo *leaf-wise growth* con límite de profundidad en vez de *Level-wise growth* que utiliza XGBoost, como se muestra en la figura 2.2.

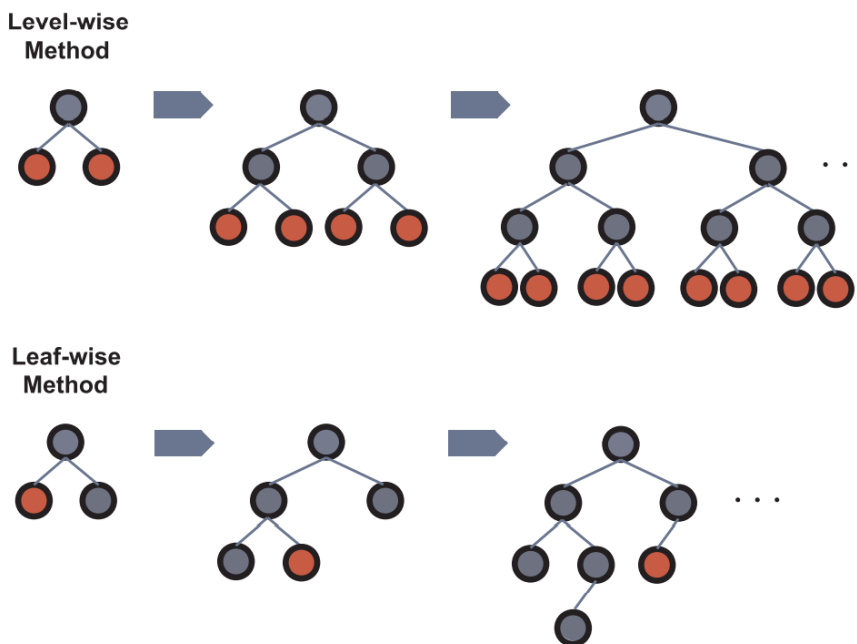


Figura 2.2: Diagrama de métodos de aprendizaje *Level-wise* y *Leaf-wise* (Fan et al., 2019).

La optimización que utiliza LightGBM se basa principalmente en 3 algoritmos, algoritmo basado en histogramas, *Gradient-based One-Side Sampling* (GOSS) y *Exclusive Feature Bundling* (EFB). En primer lugar el uso de histogramas para representar la información que consiste en discretizar en k la información en los valores propios de los vectores de características, utilizando punto flotante, reduciendo la memoria requerida en $1/8$ de la original ya que los bins se pueden guardar en enteros de 8 bits, por lo que el uso de memoria se reduce en $1/8$ con respecto a la original (Gan et al., 2021), como se observa en la figura 2.3 (A). La siguiente optimización se observa en la figura 2.3 (B) correspondiente al algoritmo GOSS (Ke et al., 2017) que consiste en una reducción de la data a entrenar, donde se toma un porcentaje $a\%$ de los valores con mayor residuos, además de tomar un porcentaje $b\%$ de la data restante ($100\% - a\%$). El algoritmo GOSS permite reducir la cantidad de datos al momento de entrenar sin afectar significativamente la precisión del algoritmo final. El algoritmo EFB permite reducir la cantidad de características en parámetros mutuamente excluyentes como se muestra en la figura 2.3 (C), donde el vector de características 1 y 2, se combinan para generar un paquete de características nuevo que no afecta la distribución de los valores de salida (target), pero si disminuye el tamaño del vector de entrada.

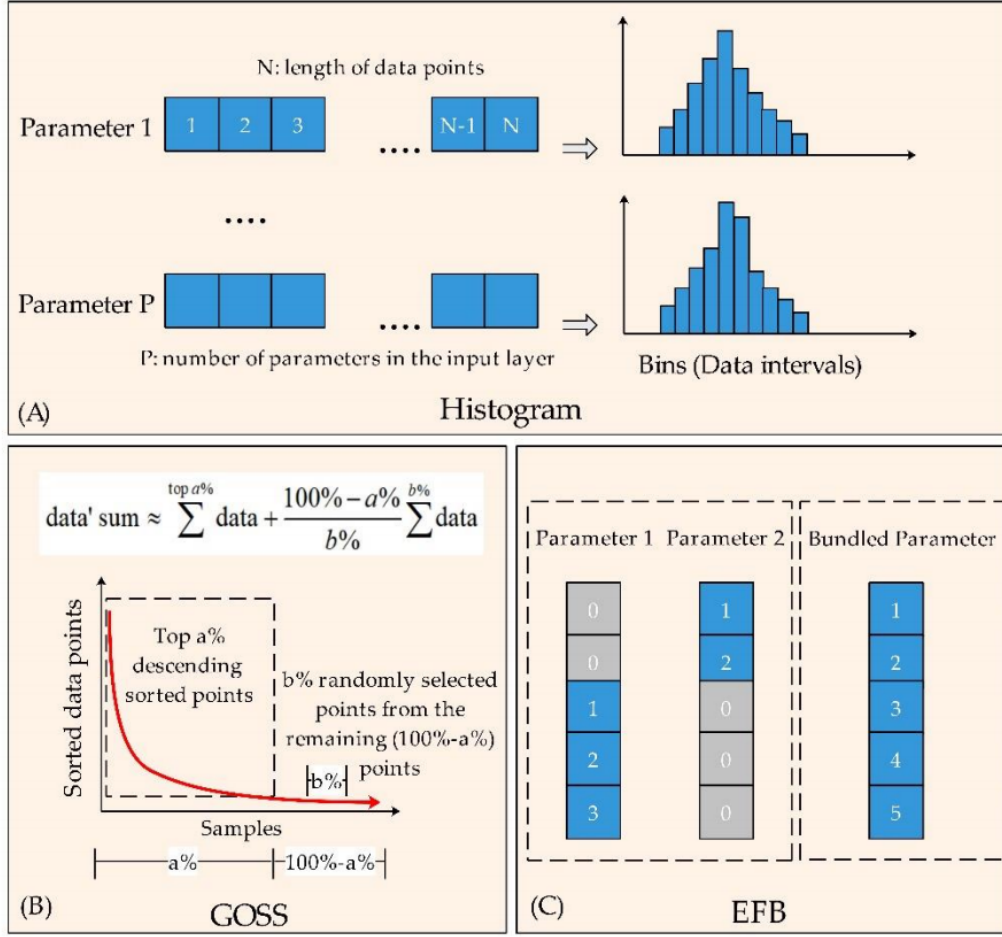


Figura 2.3: (A) Histogramas (B) Gradient-based one-side sampling (C) EFM. Diagramas de algoritmos utilizados en LightGBM para acelerar el entrenamiento. (Gan et al., 2021).

Para comprender mejor en que se basa LightGBM es que a continuación se describe el proceso matemático (Gan et al., 2021). Dado un set de entrenamiento $X = \{(x_i, y_i)\}_{i=1}^n$, se busca una aproximación $\hat{f}(x)$ a una función $f^*(x)$ que minimice la función de pérdida $L(y, f(x))$ como la siguiente expresión:

$$\hat{f} = \underset{f}{\operatorname{argmin}} E_{y,x} L(y, f(x)) \quad (2.12)$$

LightGBM integra un numero T de árboles de regresión donde la aproximación final del modelo es:

$$f_T(X) = \sum_{t=1}^T f_t(X) \quad (2.13)$$

Cada árbol de regresión es definido como $w_q(x)$, $q \in \{1, 2, \dots, N\}$, donde N representa el numero de hojas, q es la regla de decisión para cada árbol y w corresponde a el peso de cada nodo hoja. El modelo es entrenado utilizando la siguiente función objetivo que se define de forma aditiva para cada paso t:

$$\Gamma_t = \sum_{j=1}^N L(y_i, F_{t-1}(x_i) + f_t(x_i)) + \Omega(f_t(x_i)) \quad (2.14)$$

Donde Ω corresponde a una factor de regulación de aprendizaje (Gan et al., 2021), que tiene la siguiente expresión:

$$\Omega(f_t) = \alpha \sum_{j=1}^J |w_j| + \frac{1}{2} \beta \sum_{j=1}^J w_j^2 \quad (2.15)$$

Esto quiere decir que sigue la regla de minimizar la función de perdida en cada iteración, es decir, cada vez que se suma un árbol. Utilizando la aproximación de Newton.

$$\Gamma_t \cong \sum_{j=1}^N L(F_{t-1}(x_i)) + (g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)) + \Omega(f_t(x_i)) \quad (2.16)$$

Para simplificar la ecuación se elimina el termino conocido para t-1, quedando:

$$\Gamma_t \cong [\sum_{j=1}^N (g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i))] + \Omega(f_t(x_i)) \quad (2.17)$$

Donde g_i y h_i corresponde al gradiente de 1er y 2do respecto a la respuesta del árbol t-1. Por lo que si se define el subconjunto I_j de la hoja j, se tiene.

$$\Gamma_t \cong \sum_{j=1}^N [(G_j w_j + \alpha |w_j|) + \frac{1}{2} (H_j + \beta) w_j^2] \quad (2.18)$$

$$G_j = \sum_{i \in I_j} g_i \quad (2.19)$$

$$H_j = \sum_{i \in I_j} h_i \quad (2.20)$$

Donde w_j para cada hoja del árbol $t \in T$ es una variable desconocidas a resolver, por lo que se se deriva la ecuación 2.18 respecto a w_j y se iguala a cero, se obtienen las siguientes expresiones:

$$w_j^* = -sgn(G_j) \frac{G_{\alpha,j}}{H_j + \beta} \quad (2.21)$$

$$G_{\alpha,j} = max\{0, (|G_j| - \alpha)\} \quad (2.22)$$

Donde sgn corresponde a la función signo. Por lo que determinando w_j^* la ecuación 2.18 se puede reescribir, obteniendo

$$\Gamma_t = -\frac{1}{2} \sum_{j=1}^J \left(\frac{G_{\alpha,j}^2}{H_j + \beta} \right) \quad (2.23)$$

De acuerdo a la ecuación 2.23, se tiene que la contribución a la reducción del error para cada nodo hoja se define como:

$$V_j = \left(\frac{G_{\alpha,j}^2}{H_j + \beta} \right) \quad (2.24)$$

Por lo que si se generan dos nuevos nodos hoja a partir del nodo j -ésimo, se obtiene la diferencia V_j después de división, la cual se define como:

$$V_d = \left(\frac{G_{L,\alpha,j}^2}{H_{L,j} + \beta} + \frac{G_{R,\alpha,j}^2}{H_{R,j} + \beta} \right) - \frac{G_{\alpha,j}^2}{H_j + \beta} \quad (2.25)$$

Donde $G_{L,\alpha,j}$, $H_{R,j}$, $G_{R,\alpha,j}$ y $H_{R,j} + \beta$ corresponden a los valores G y H de los nuevos nodos (izquierdo y derecho) a partir del nodo j -ésimo. A partir de la ecuación 2.25 se puede inferir que el modelo LightGBM, elige el parámetro q (regla de partición) que maximiza dicha ecuación, es decir, donde V_d se maximiza.

2.9. DeepLearning

La base del DeepLearning corresponde a un subconjunto del aprendizaje de máquina, donde la red está compuesta por múltiples capas y aprende de una gran cantidad de datos. La estructura general de una red neuronal se observa en la figura 2.4, donde los pesos para la primera capa para un vector de entrada p de variables $X = (X_1, X_2, \dots, X_p)$ (Daniela et al., 2021), se rigen por la siguiente ecuación.

$$A_k^{(1)} = h_k^{(1)}(X) = g(w_{k0}^{(1)} + \sum_{j=1}^p w_{kj}^{(1)} X_j) \quad (2.26)$$

Para $k = 1, \dots, K_1$. La segunda capa se inicia en la activación de la capa anterior $A_k^{(1)}$.

$$A_l^{(2)} = h_l^{(2)}(X) = g(w_{l0}^{(2)} + \sum_{k=1}^{K_1} w_{lk}^{(1)} A_k^{(1)}) \quad (2.27)$$

Para $l = 1, \dots, K_2$.

2.10. Redes Recurrentes

Debido a que una gran cantidad de data es naturalmente secuencial, es que surge la necesidad de integrar información del tiempo pasado (predicho por el algoritmo) en el diseño de la red, las cuales reciben el nombre de Redes Neuronales Recurrentes (RNN por sus siglas en ingles).

Como se observa en la figura 2.5, la secuencia de entrada X se procesa de manera secuencial y cada elemento X_l alimenta la cada oculta A_l y a su vez está tiene como entra la el vector de activación A_{l-1} , sin embargo, la salida Y solo depende directamente de A_L (Daniela et al., 2021). Estas redes son comúnmente utilizadas en:

- Procesado de series de temperatura, lluvia, velocidad del viento y calidad de aire.
- Finanzas, con el fin de seguir patrones de marcado.
- Procesamiento de voz, grabación de música para realiza transcripciones.
- Análisis de escritura a mano.

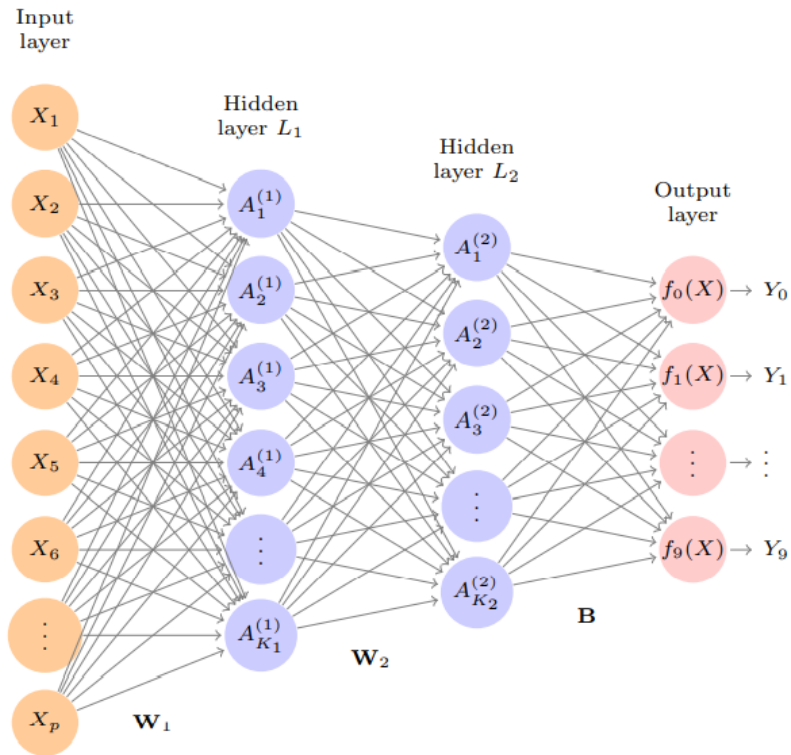


Figura 2.4: Diagrama de Red Neuronal con 2 capas ocultas y múltiples salidas, utilizado en el problema de *MNIST handwritten-digit*, la entrada tiene 783 unidades, las siguientes dos capas tienen $K_1 = 256$ y $K_2 = 128$ unidades respectivamente (Daniela et al., 2021).

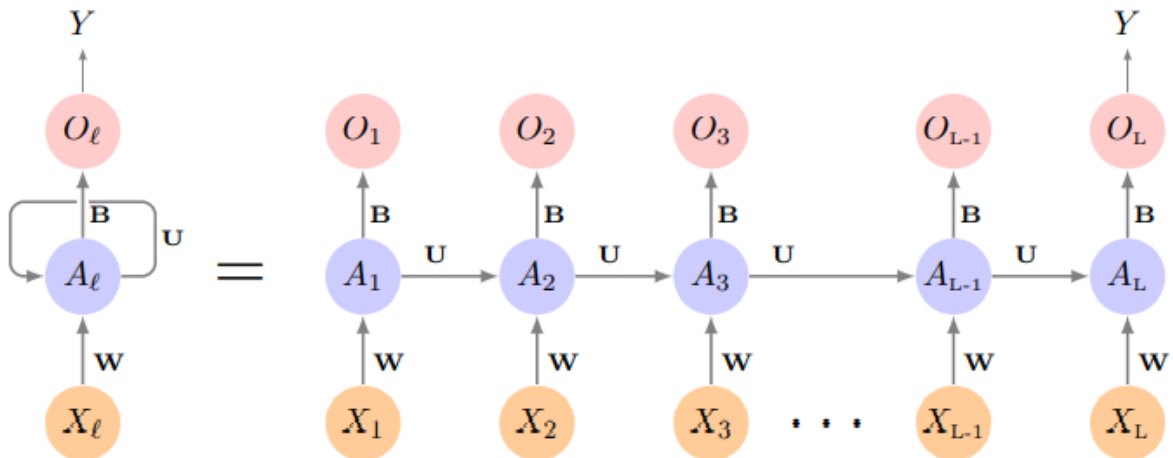


Figura 2.5: Diagrama de una RNN simple, donde la secuencia de entrada es $\{X_t\}_1^L$ y la salida es única (Daniela et al., 2021).

2.11. Long Short-Term Memory Recurrent Neural Network (LSTM)

LSTM corresponde a una arquitectura de red artificial recurrente (RNN), utilizada en Deep Learning (DL). La cual se compone de unidades especiales llamada bloques de memoria, con autoconexiones que almacenan estados temporales de la red. Las principales unidades de la unidad LSTM esta compuesta por una *cell (C)*, *input gate(IG)*, *output gate (FG)* y *forget gate(OG)* (Graves et al., 2009). Donde los elementos *gates* controlan el flujo de información a la celda (*cell*) de memoria como se observa en la figura 2.6.

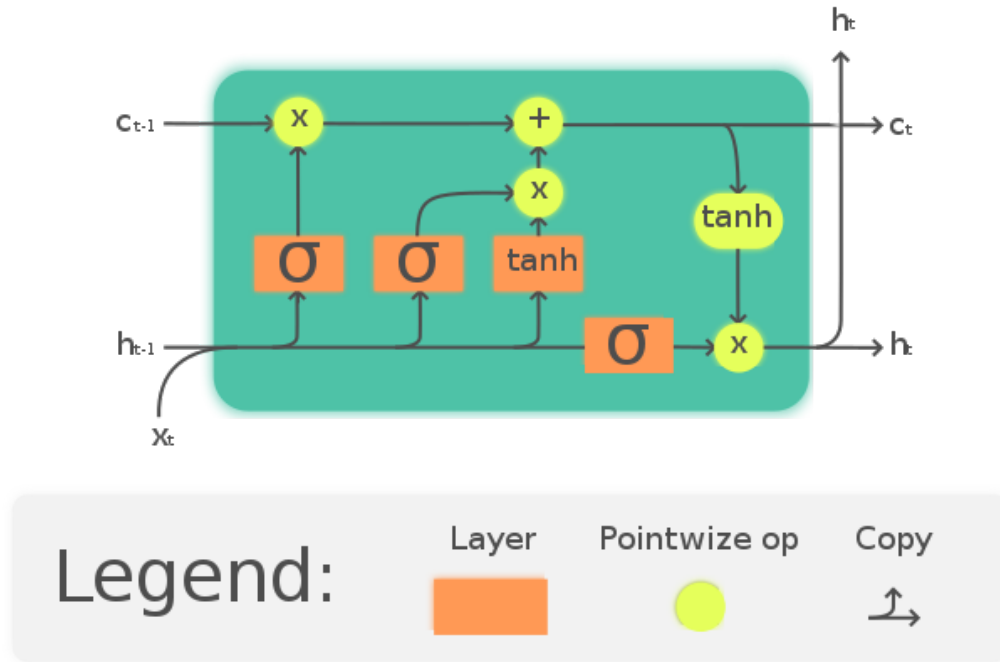


Figura 2.6: Diagrama de una celda LSTM (Chevalier, 2018).

LA red LSTM mapea la secuencia de entrada $x = \{x_1, \dots, x_T\}$ a una secuencia de salida $y = \{y_1, \dots, y_T\}$, donde para cada unidad de la red se calculan las siguientes ecuaciones de $t = 1$ a T :

$$i_t = \sigma(W_{ix}x_t + W_{im}m_{t-1} + W_{ic}c_{t-1} + b_i) \quad (2.28)$$

$$f_t = \sigma(W_{fx}x_t + W_{fm}m_{t-1} + W_{fc}c_{t-1} + b_f) \quad (2.29)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g(W_{cx}x_t + W_{cm}m_{t-1} + b_c) \quad (2.30)$$

$$o_t = \sigma(W_{ox}x_t + W_{om}m_{t-1} + W_{oc}c_{t-1} + b_o) \quad (2.31)$$

$$m_t = o_t \odot h(c_t) \quad (2.32)$$

$$y_t = \phi(W_{ym}m_t + b_y) \quad (2.33)$$

Donde los términos W corresponde a los pesos de entrada a las puertas de IG (W_{ix}), OG (W_{ic}), FG (W_{ix}) y los pesos W_{ic} , W_{fc} , W_{oc} , correspondiente a una matriz diagonal de pesos para la *peephole connections*, que permite no solo obtener tomar información de salida previa, si no tambien el estado previo de al celda ponderada por la salida de las puertas. Los términos b correspondes a los *bias*, σ es la función lógica *sigmoid*. A la salida del sistema se realiza producto punto con un vector de activación m y finalmente h y g corresponde a funciones de activación de C (Chevalier, 2018).

2.12. Incertidumbre en Red LSTM

Los modelos basados en *Deep Learning* tienen cada vez más aplicaciones en diversos campos, por lo que resulta importante lograr representar la incertidumbre del modelo en aplicaciones donde esto es crucial. Muchas veces se confunde la salida aplicando *softmax*, sin embargo, incluso modelos con esta herramienta presenta errores en su salida. Es por esto que se decide utilizar la herramienta *Montecarlo Dropout*, la cual consistente en desactivar nodos de las capas ocultas del modelo. (Gal y Ghahramani, 2016)

La herramienta Dropout se utiliza ampliamente con el fin de evitar sobreajuste en el entrenamiento, sin embargo, tambien es útil para captar la incertidumbre en la etapa de predicción, desactivando nodos aleatoriamente (Abdar et al., 2021)(Gal y Ghahramani, 2016). Si se repite la predicción para una mismo vector de entrada $x_i = \{x_1^i, \dots, x_N^i\}$ se obtiene una salida $y_i = \{y_1^i, \dots, y_N^i\}$, donde i corresponde a la dimensión del vector de entrada y N la cantidad de repeticiones de estimación. Si N es lo suficientemente grande se logra obtener la distribución asociada al espacio de salida para x_i . Un ejemplo de esto se muestra en la figura 2.7, donde se analiza la esta aplicación en regeneración 3D de resonancias magnéticas (MRI).

2.13. SHapley additive exPlanations (SHAP)

Es un concepto de la teoría de juegos, desarrollado originalmente como una medida para distribuir equitativamente una recompensa entre un conjunto de jugadores que contribuyen a un determinado resultado (Shapley, 1952). En el contexto de aprendizaje de máquina la entrada los jugadores son las características de entrada y el resultado es la decisión del modelo (Lundberg y Lee, 2017). SHAP crea entradas simplificadas z asignando x a z a través de $x = h_x(z)$. Basado en z , el modelo original $f(x)$ se puede aproximar con una función lineal:

$$f(x) = g(z) = \varphi_0 \sum_{i=1}^M \varphi_i z_i \quad (2.34)$$

Donde $z = 0, 1^M$, M es el número de características de entrada, $\varphi_0 = f(h_x(0))$, y φ_i es el valor de contribución de las características:

$$\varphi_i = \sum_{S \in F \setminus \{i\}} \frac{|S|!(M - |S| - 1)!}{M} [f_x(S \cup \{i\}) - f_x(S)] \quad (2.35)$$

$$f_x(S) = f(h_x^{-1}(z)) = E[f(x)|x_s] \quad (2.36)$$

Donde S es un subconjunto de F , siendo F un conjunto no nulo en z , es decir el conjunto

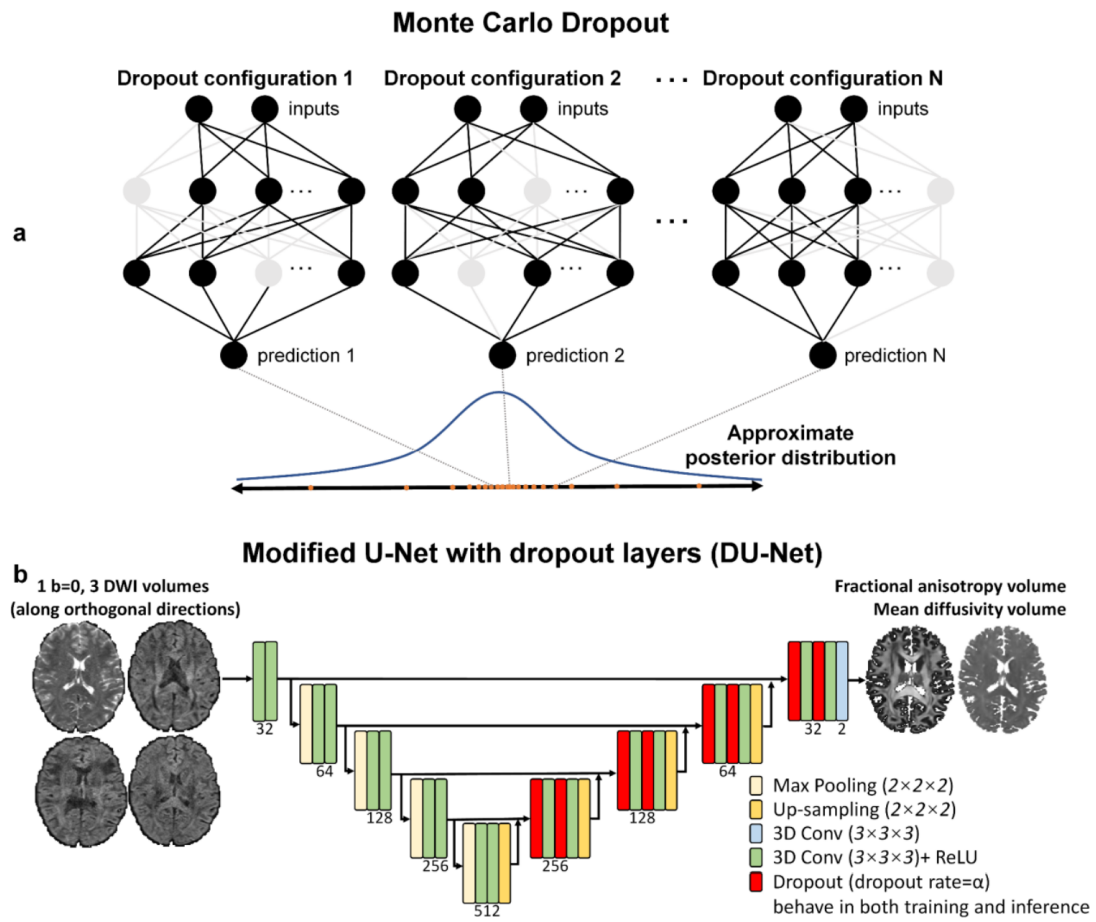


Figura 2.7: Ejemplo de red U-Net 3D con aplicación de *dropout* en las capas de decodificación, tanto en el entrenamiento como en la validación. (Avci et al., 2021)

de características. El termino de la derecha en la ecuación 2.35 es la contribución marginal. Siendo esta última la variable más desafiante de estimar, donde la librería *SHAP values* de Python ofrece múltiples aproximaciones dependiendo el algoritmo utilizado, es este caso se utiliza *TreeSHAP* y *DeepSHAP* (Edoardo, Ferenc, Stella, Daniel, y Georg, 2022).

Capítulo 3

Estado del Arte

En el presente trabajo se busca generar un modelo (o serie de modelos) que sea capaz de estimar el consumo de un EV dado una ruta, utilizando herramientas de *Machine Learning*. Bajo esta premisa existen diversos trabajos que buscan estimar el consumo de un EV, algunos como un indicador virtual, que a partir del modelo mecánico estima el consumo, otros tienen el mismo objetivo pero utilizando data para crear un modelo caja negra. Así también, existen modelos más completos basados en la fenomenología, que buscan estimar tanto la velocidad del vehículo en un ruta específica como el consumo a lo largo de esta. Por otro lado, existen trabajos que se basan en la fenomenología, pero utilizan data histórica para entrenar sistemas de multi-modelos que permiten estimar parámetros relacionados con la cinética del vehículo, para luego realizar estimar el consumo eléctrico total de la ruta. A pesar de los trabajos previamente realizados, hasta la fecha no existía una base de datos a gran escala que permitiera utilizar herramientas de *Machine Learning* más actualizadas. Gracias a la base de datos proporcionada por G. S. Oh es que se tiene la opción de realizar un estudio de la estimación de consumo basado en data, proponiendo una estimación de rango variable con la finalidad de optimizar los resultados, en donde se puede tener una estimación dentro del rango de 50 a 150 metros así como la estimación del consumo en la ruta completa, utilizando herramientas que permiten definir el grado de incertidumbre asociado a cada estimación. Según a lo anteriormente mencionado, es que puede ver que existen trabajos previos con el que se puede comparar el presente trabajo, sin embargo, existen aplicaciones en el que el modelo construido presenta ventajas no comparables con trabajos previos.

Según la revisión bibliográfica realizada por Yuche Chen, los trabajos realizados en torno a la estimación de consumo de un EV, se puede separar en dos grandes metodologías, como se observa en la figure 3.1 ,la primera llamada *Rule-based* que consiste en trabajos que se basan en la fenomenología interna del vehículo y así crear sub-modelos que reflejan el funcionamiento mecánico y la interacción entre ellos, la ventajas de estos modelos es que son fácilmente interpretables, sin embargo, la precisión de estos modelos no es alta y teniendo aun desafíos que superar para su aplicación a nivel de flota de vehículos (Y. Chen et al., 2020). Por otro lado tenemos la llamada *Data-driven* que consiste en generar modelos basado en data históricas, en donde la ventajas de estos modelos es que permiten la integración de múltiples variables de forma sencilla, además de que permiten utilizar las últimas herramientas asociadas a *Machine Learning* como redes neuronales, clustering, entre otros, la desventaja es que son menos interpretables (Y. Chen et al., 2020).

A continuación, se presentan algunos de los trabajos más relevantes en el desarrollo de modelos de estimación de consumo de EV.

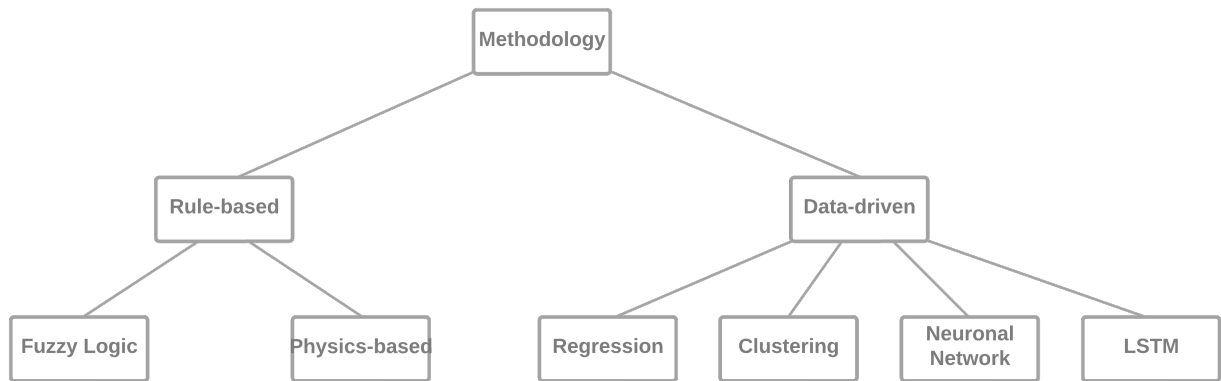


Figura 3.1: Clasificación del estado del arte de estimación de consumo en EV, adaptado de (Y. Chen et al., 2020).

3.1. Determinación de los principales factores que influyen en el consumo de vehículos eléctricos

Como se muestra en el estudio realizado por Wne Li, el consumo de energía de un EV depende de múltiples factores, como se observa en la figura 3.2, donde se ve que según la literatura estos se agrupan en 6 grandes factores (Li et al., 2016), los cuales se definen a continuación.

- **Tecnología y vehículo:** El sistema de baterías es de los componentes más importantes asociados al consumo. Factores como el número de celdas, tipo de baterías y sistema de gestión de baterías, afectan directamente la capacidad de batería, energía máxima y la masa. Por otro lado factores asociados al desgaste del paquete de baterías que se define como SOH. Otro punto importante es la capacidad de carga regenerativa que no tiene efecto en nivel de SOC elevados o altas temperaturas del paquete de baterías.

Componentes auxiliares como el sistema de HVAC que controla la temperatura del paquete de baterías y la interna del vehículo, juega un papel considerable en el consumo que se requiere altas cantidades de energía. Otros componentes auxiliares como la radio, sistemas de navegación, entre otros tiene un impacto en el consumo en menor medida.

- **Condiciones Ambientales Artificiales :** Estos factores incluyen la infraestructura , nivel de urbanización, interacciones con el peatón y demás automovilistas que se refleja en las intersecciones, tráfico, señales y luces de tránsito. Otros factores corresponden a la gestión del tráfico, y condición de la ruta.
- **Condiciones Ambientales Naturales:** Los factores generales asociados a este grupo corresponden a la topografía, clima, entre muchos otros. En lo que refiere a la topografía se tienen factores como la inclinación de la ruta, tanto en ascenso como en descenso,

que tienen impacto en el consumo (ascenso) o regeneración (descenso). Por otro lado los factores asociados al clima corresponden a la estación climática, lluvia, nieve, visibilidad de la ruta, hora del día, entre otros. Afectando la forma de conducción y por ende impactando directamente en el consumo.

- **Conducción:** El factor que define la conducción de define mediante al agresividad, que define los patrones de aceleración y desaceleración. Mientras más agresivo el estilo de conducción, mayor son los patrones de aceleración y desaceleración, lo que lleva a un mayor consumo. Estos patrones están determinados por diversos factores, entre ellos, edad, género, nivel de confort dentro del automóvil (condiciona uso de auxiliares), factores físicos y psicológicos, entre otros.
- **Tipo de viaje:** El tipo de viaje y la intenciones de este, determinarán factores como la frecuencia con que se recorre una ruta, duración del viaje, urgencia o necesidad del viaje. Por ejemplo los patrones de conducción son distintos en vacaciones.
- **Métricas:** Los factores asociados a las métricas que se obtienen en los distintos estudios asociados el consumo eléctrico de un EV. Por ejemplo estudios muestran que cuando las baterías están más frías el consumo es mayor. Otros factores que tienen relación en este punto es la caracterización del paquete de baterías, como el SOC. También se definen parámetros como la experiencia de la persona que conduce a la hora de realizar las pruebas, la precisión al obtener la data, hardware y software y diseño del experimento.

En dicho trabajo se busca ver el impacto de distintos factores en el consumo, en donde el experimento es diseñado utilizando un EV Nissan Leaf 2011, conducido sin ningún tripulante más que el conductor por el área de RandWick, Sydney. Esta ruta tiene un largo de 4.17km y se realizaron 25 vueltas.

Dado que estudiar como impactan todas los factores señalados en al figura 3.2, es que solo se toman 4 de estos, correspondiente a la topografía, infraestructura, trafico y clima. De los cuales se concluye que los parámetros más importantes son la topografía y el clima, el primero tiene un mayor impacto debido a que los ascensos y descensos favorecen al consumo o regeneración. En cuanto al clima se observa que el uso de HVAC puede variar la autonomía del EV de 100 a 222 km.

3.2. Fenomenología del Consumo en Vehículos Eléctricos

El punto de partida para ver los parámetros necesarios para realizar un modelo basado en data, es entender la fenomenología del consumo de un EV. En la figura 3.3 se observa las fuerzas que interactúan en el movimiento de un EV, donde F_a , F_{air} , F_c , F_r corresponden respectivamente a la fuerza de aceleración, pendiente, resistencia aerodinámica y de roce. Siendo la potencia la suma de estas fuerzas multiplicada por la velocidad representada en la ecuación 3.1.

$$P_{total} = F_a \cdot V + F_{air} \cdot V + F_c \cdot V + F_r \cdot V = m \cdot a \cdot V + \frac{1}{2} \cdot \rho_{air} \cdot C_d \cdot A \cdot V^3 + m \cdot g \cdot \sin(\theta) \cdot V + m \cdot g \cdot K_r \cdot V \quad (3.1)$$

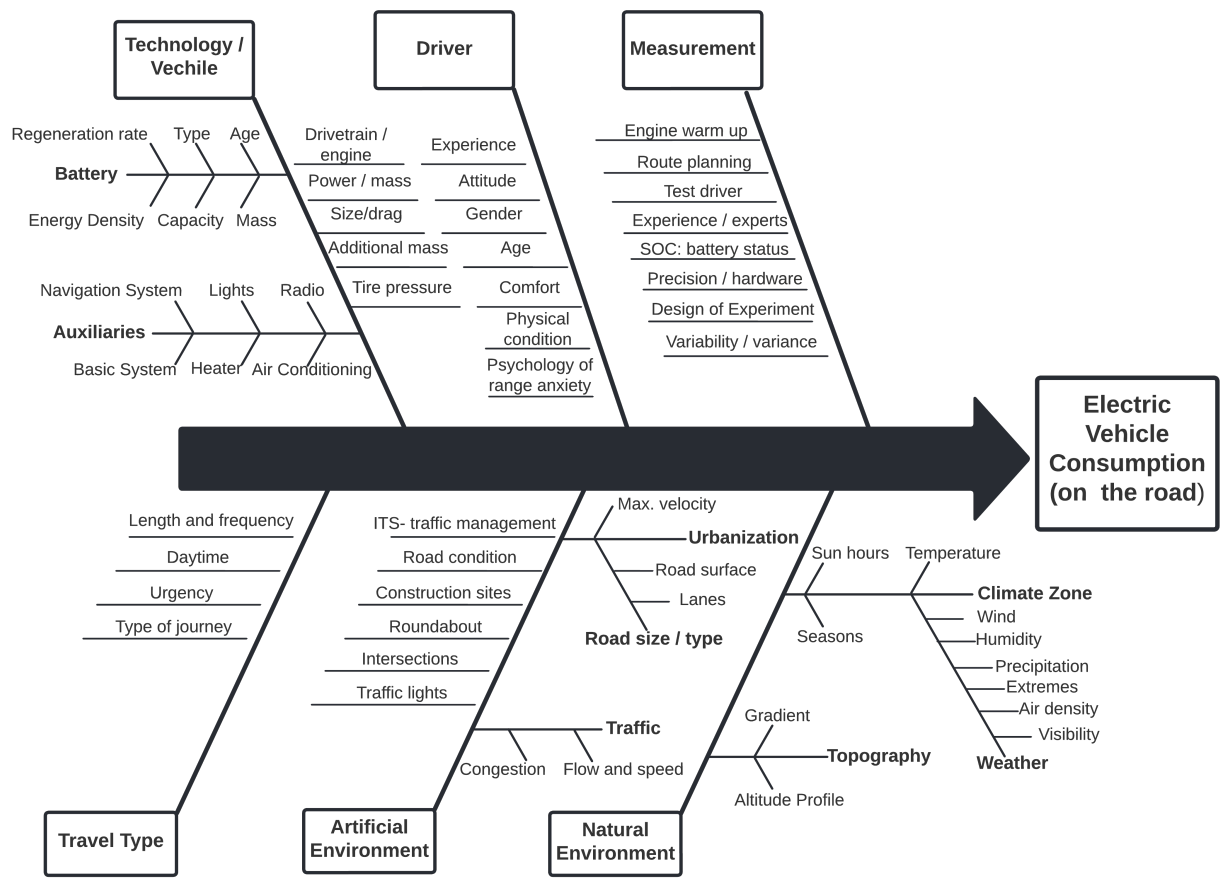


Figura 3.2: Diagrama Ishikawa de los factores que impactan en el consumo eléctrico de un EV, adaptado de (Li et al., 2016).

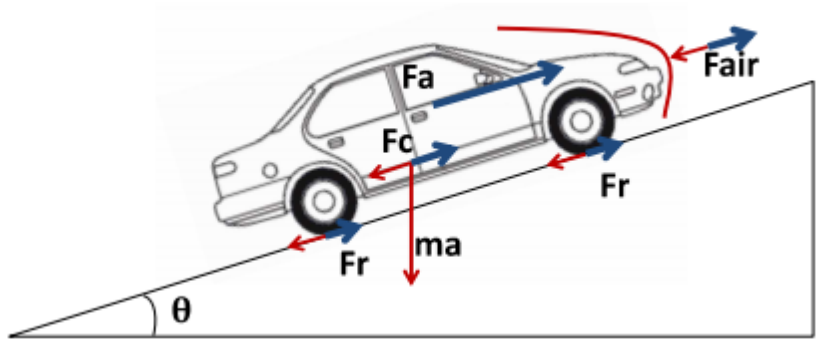


Figura 3.3: Fuerzas involucradas en el requerimientos de potencia (Kim et al., 2013).

Tabla 3.1: Caracterización de los parámetros de requerimiento de potencia.
Adaptado de (Kim et al., 2013)

	a [m/s^2]	V [m/s]	m [Kg]	g [m/m^2]	K_r
Descripción	Aceleración	Velocidad	Masa	Cte Aceleración	Coef. Roce
Dinámica	Muy alto	Alto	Muy Bajo	Muy Bajo	Alto
Dependencia	Conducción, Camino, Trafico	Conducción, Camino, Trafico	Estructura del vehículo	Altitud	Llanta, Camino

	α [°]	ρ_{air} [Kg/ m^3]	C_d	A [m^2]
Descripción	Inclinación	Densidad Aire	Coef. Arrastre	Area Frontal
Dinámica	Alto	Bajo	Muy Bajo	Muy Bajo
Dependencia	Camino	Altitud	Estructura del Vehículo	Estructura del Vehículo

Los factores de las ecuaciones de fuerzas involucradas en el consumo de un vehículo se detalla en la tabla 3.1, donde estos parámetros se pueden dividir en 3 categorías.

- **Estable:** Corresponde a los parámetros que tienen baja variabilidad y corresponde a masa del vehículo (m) , Coeficiente de arrastre(C_d), Área frontal (A), densidad de Aire (ρ_{air}), aceleración de gravedad (g).
- **Semi-Dinámico:** Corresponde a los parámetros que tienen Variabilidad pero depende de las condiciones de la ruta, donde se tiene la inclinación de la ruta (θ) , Coeficiente de roca (K_r).
- **Dinámico:** Estos parámetros tienen alta variabilidad y depende de variables exógenas difíciles de estimar, corresponde a aceleración (a) y velocidad del vehículo (V).

3.2.1. Métodos de estimación de aceleración

Los parámetros estables y semi-dinámicos se pueden estimar utilizando análisis de los parámetros intrínsecos de la ruta, sin embargo, los parámetros altamente dinámicos como la aceleración y la velocidad necesitan de una estimación online, por lo que en este paper se proponen varias formas. entre las que se encuentran:

1. **Aceleración Previa:** Este método utiliza la aceleración en presente para estimar la aceleración futura. Esta metodología es efectiva cuando existen baja fluctuación en la aceleración y una alta tasa de muestreo.
2. **Aceleración previa con fuerza de reacción:** Se busca estimar la aceleración futura utilizando la pendiente de aceleración en $n-1$ y n , ajustando el cambio de velocidad para

n , utilizando la ecuación 3.2

$$a_{n+1}^{\hat{}} = a_n + k(a_n - a_{n-1}) \quad (3.2)$$

donde el parámetro k corresponde a un coeficiente que ajuste la sensibilidad de la estimación de aceleración a la cambio en los tiempos anteriores.

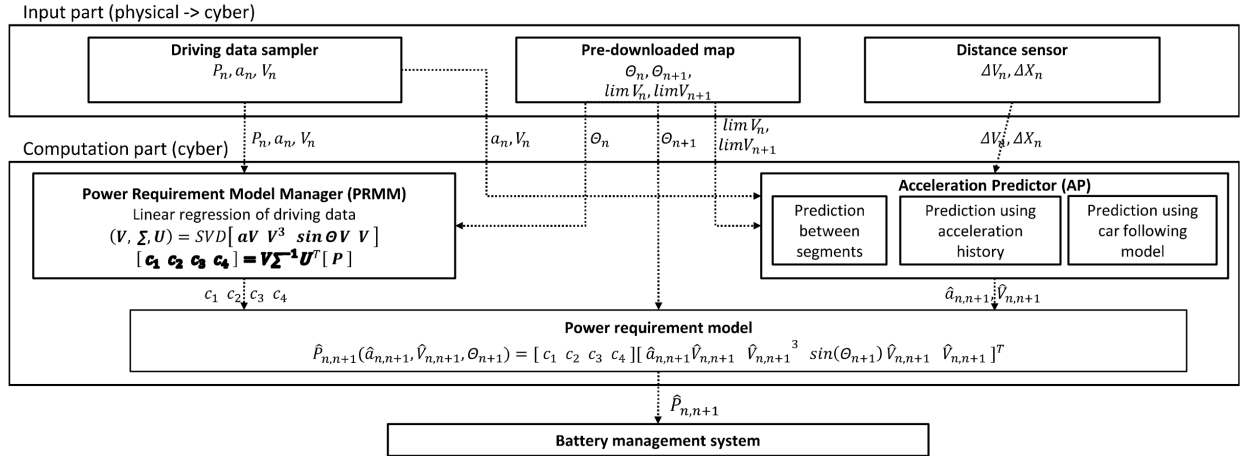
3. **Correlación con aceleraciones pasadas:** Dado que existe una alta correlación entre aceleraciones pasadas, se realiza un filtro lineal en donde se toma el error asociado entre la estimación y el valor medido, luego se realiza el producto interno entre este error y el vector de correlaciones. Este método es útil cuando existe una alta correlación entre la aceleración futura y las aceleraciones históricas cercanas, pero requiere mayor poder de cómputo que los métodos anteriores.
4. **Modelo de seguimiento de automóvil:** Muchos estudios de tráfico estudian los comportamientos microscópicos del conductor para analizar el flujo de tráfico en tiempo real. Este método describe la aceleración a través de la velocidad relativa entre el automóvil a estudiar y el vehículo del frente.
5. **Segmentación de ruta:** Se realiza una segmentación de la ruta basado en las regulaciones del tráfico, de esta forma se estima la aceleración cuando el automóvil pasa de un segmento a otro. Debido a que esta información puede ser cargada con anterioridad, es que no tiene dependencia con la velocidad de muestreo.

3.2.1.1. Evaluación

Para realizar la evaluación de los distintos modelos de estimación es que se realizan distintas pruebas, donde se definen como:

- Prev (Aceleración previa)
- Prev+ (Aceleración previa con fuerza de reacción)
- Corr (Correlación con aceleraciones pasadas)
- CF (Seguimiento de automóvil)
- RS (Segmentación de Ruta)
- Uni(H) (unificación de método de estimación basado Seguimiento de automóvil y metodología basada en datos históricos. **H** puede ser Prev, Prev+ o Corr)

En estas pruebas se realizan utilizando el PRP, mostrado en la figura 3.4, con frecuencias de muestreo de 2, 5 y 10 Hz, donde se asume que el efecto de frenado regenerativo no existe.



<Notations> P_n, a_n, V_n : power demand, acceleration, speed *measured* at step n
 $\hat{P}_{n,n+1}, \hat{a}_{n,n+1}, \hat{V}_{n,n+1}$: power demand, acceleration, speed at step $n + 1$ *estimated* at step n
 $\theta_n, \lim V_n$: road slope, speed limit at step n obtained from a pre-downloaded map
 $\Delta V_n, \Delta X_n$: relative speed and distance between the leading and my cars *measured* at step n

Figura 3.4: Estructura de sistema de predicci3n de requerimiento de potencia (PRP), extraído de (Kim et al., 2013)

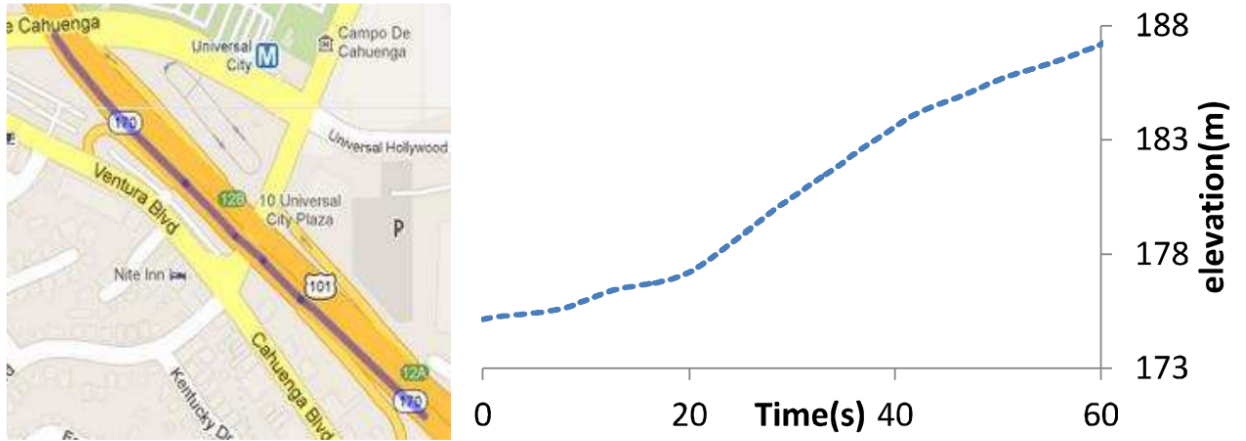


Figura 3.5: La ruta de autopista US-101, Los Angeles, USA, en direcci3n sur. Extraído de (Kim et al., 2013)

Los resultados obtenidos para la diferentes metodologías , se comparan con el al error inherente 3ptimo(Opt-P), que corresponde al error asociado a una estimaci3n perfecta del consumo de potencia pero que tiene un error asociado por el hecho de tener una estimaci3n discreta a lo largo del tiempo, como se muestra en la figura 3.6. Por otro lado lo mismo sucede con el error inherente de la aceleraci3n (Opt-A) el cual corresponde al m3nimo error al que se puede aspirar debido al sub-muestreo asociado a la metodolog3a.

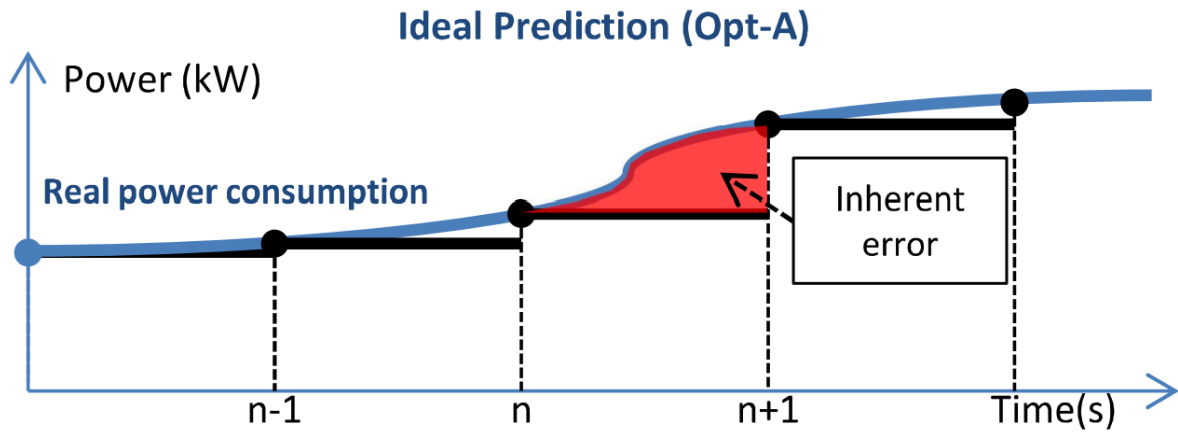


Figura 3.6: Error inherente por estimación discreta, extraído de (Kim et al., 2013)

Los resultados obtenidos se presentan en la tabla 3.2, donde se puede ver como la combinación entre segmentación de ruta, seguimiento de automóvil y correlación de aceleración de data histórica, corresponde al mejor resultado para la ruta estudiada.

Tabla 3.2: Resultados de metodología implementada, adaptado de (Kim et al., 2013)

Métodos de predicción	Error promedio kW		
	2Hz	5Hz	10Hz
Opt-P	8.671	3.861	1.705
PRP con Opt-A	9.186	4.448	2.458
Potencia actual Previa	15.884	10.461	5.904
PRP con Prev +	17.198	11.212	5.202
PRP con Corr	12.567	10.368	4.962
PRP con CF	10.464	11.006	10.630
PRP con Uni(Prev+)	12.138	10.425	4.916
PRP con Uni(Corr)	10.891	10.220	5.068

De esta metodología se puede ver una comprensión de problema donde se segmenta la ruta por secciones definidas por restricciones del tránsito es beneficioso, además de que queda en evidencia la causalidad de este problema, donde la velocidad y aceleración dependen del instante anterior. Por otro lado una división del problema donde parte del estudio se enfoca en el estudio de la energía y otro en un desarrollo de modelo de estimación de los parámetros más importantes resulta en un metodología que permite el desarrollo de problema de forma clara, con resultados concluyentes.

3.2.2. Soluciones basadas en Machine Learnig

Como se ha visto en distintos estudios, la implementación de modelación a través de metodología basada en *Machine Learning* (ML) ha ido en aumento y esto tambien se observa en los artículos que buscan estimar el consumo en energía en vehículos eléctricos. A continuación se presentan dos artículos que en se metodología utilizan información del ambiente natural como del artificial, para estimar la velocidad y luego entrenar un modelo neuronal para obtener el consumo de energético.

3.2.2.1. *Energy consumption estimation in electric vehicles considering driving style*

En este articulo se observa que se realiza el estudio de (%Bateria / Km) de EV utilizando un automóvil Mitsubishi i-Miev, recorriendo 3 rutas por distintos conductores, utilizando 2 para training del modelo y la última como validación, donde la data se procesa cada 5 minutos, para realizar la ingesta al modelo.

Donde se toman múltiples variables que se dividen en 2 grandes grupos, relacionadas al comportamiento del conductor (velocidad, cambios de velocidad, etc) y otras relacionadas a restricciones de la ruta (Límites de velocidad, inclinación, etc). Luego de esto se entrena un modelo *Multilayer Perceptron* (MLP) con distintas estructuras obteniendo los siguientes resultados.

Tabla 3.3: Resultados de metodología utilizada, adaptado de (Felipe et al., 2015)

Driving Style	Calm	Agressive
Real Consumption (% battery charge/km)	0.7034	0.938
NN	Errors is predicted consumption	
137-1	-0.0103 -4.52 %	-0.0165 -5.42 %
43-1	+0.0331 +14.44 %	+0.0590 +19.32 %

3.2.2.2. *A Data-Driven Method for Energy Consumption Prediction and Energy-Efficient Routing of Electric Vehicles in Real-World Conditions*

En este estudio realizado en la ciudad de Bruselas, Belgica, se utiliza un automovil Ford Connect que fue transformado a EV, el cual fue conducido por 3700 km (Dataset 1) y tambien dos vehículo Nissan Leaf (2014) que fueron conducidos por 10700 km en total (Dataset 2).

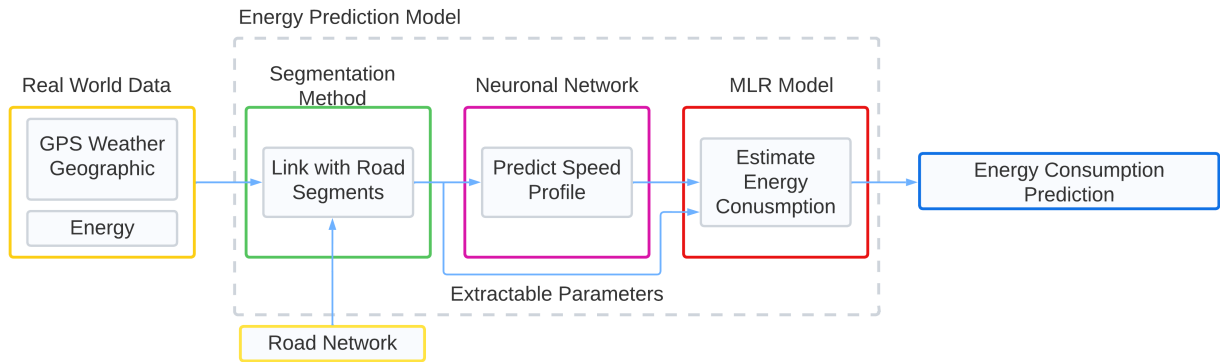


Figura 3.7: Diagrama de solución propuesta, adaptado de (Cauwer et al., 2017)

Utilizando la data adquirida, se plantea una solución como la que se presenta en el diagrama 3.7, donde se obtiene los datos en línea, sumado a datos de la ruta para estimar los perfiles de velocidad utilizando una red neuronal MLP, con la salida de esta red se ajusta una función basada en la fenomenología, *Multiple Linear Regression* (MLR), del problema para luego estimar el consumo de energía. Esta metodología se compara con la estimación utilizando una red neuronal en vez de MLR y con una estimación utilizando la multiplicación de la distancia recorrida por el consumo histórico, obteniendo los siguientes resultados.

Tabla 3.4: Resultados obtenidos, adaptado de (Cauwer et al., 2017)

Dataset	Performance Indicator	NN-MLR	NN Prediction	Average Consumption Prediction	$\langle E \rangle$ (kWh)
Dataset 1	RMSE	0.605	0.539	0.471	2.4
	MAE	0.335	0.364	0.316	
	MAE/ $\langle E \rangle$	0.14	0.15	0.13	
Dataset 2	RMSE	0.142	0.178	0.178	0.78
	MAE	0.0917	0.119	0.119	
	MAE/ $\langle E \rangle$	0.12	0.15	0.15	

3.2.2.3. *Energy consumption analysis and prediction of electric vehicles based on real-world driving data*

En este estudio realizado con data de recolectada por *National Monitoring and Management Platform for New Energy Vehicles* (NEVs) en China, donde se incluyen taxis eléctricos, buses eléctricos y vehículos de sanitización eléctricos. Donde se realizan distintos análisis con el fin de caracterizar las distribuciones de los distintos requerimientos eléctricos, con el fin de crear un modelo Markov Monte Carlo (MMC) para predecir las condiciones de conducción (DC) futuras, donde se estima la velocidad asociada a cada DC y luego de realiza una estimación de la velocidad utilizando un modelo un árbol de regresión XGBoost. El diagrama completo del framework propuesto se presenta en la figura 3.8.

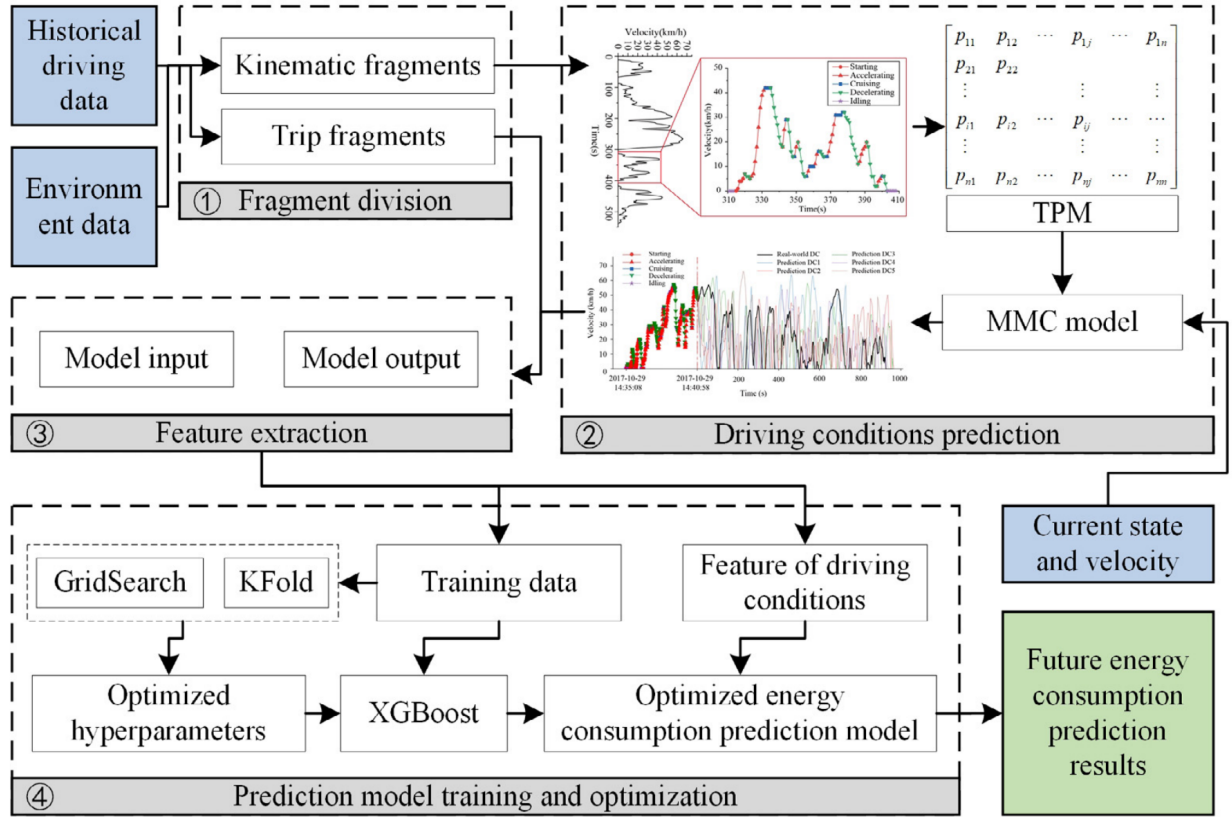


Figura 3.8: Diagrama de framework basado en ML para estimación de energía(Zhang et al., 2020)

El resultado de esta metodología se compara con un modelo convencional definido por:

$$EC_b^{pred}(x) = ECR_b^{pred}(x)M_{future}(x) \approx \frac{EC_b^{past}(x)}{M_{past}(x)}M_{future}(x) \quad (3.3)$$

Donde $M_{future}(x)$ corresponde a la distancia restante desde la posición z ECR_b^{pred} corresponde al consumo de energía estimado para la ruta restante M_{future} ; $M_{past}(x)$ utilizado en la estimación de $ECR_b^{past}(x)$. $EC_b^{past}(x)$ corresponde a la energía consumida durante la distancia recorrida $M_{past}(x)$.

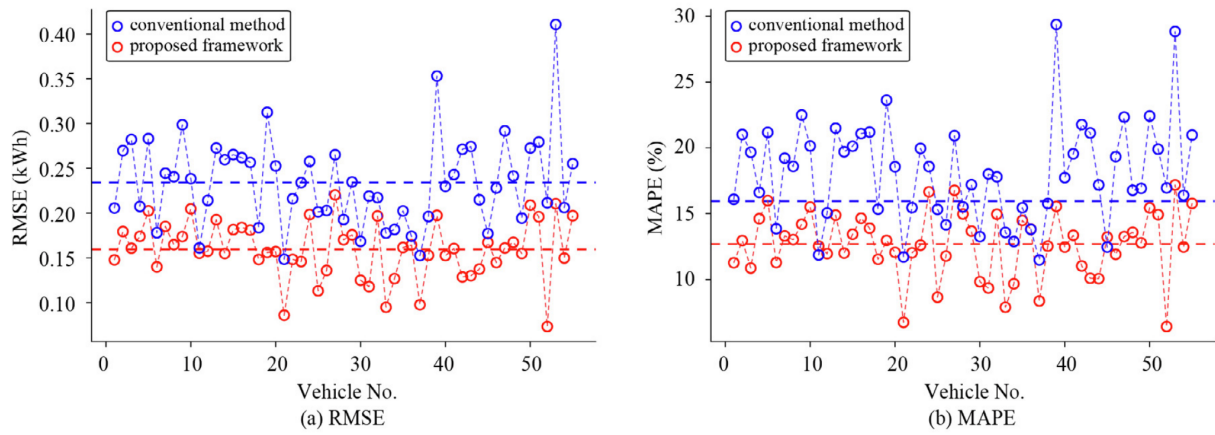


Figura 3.9: Resultados del RMSE y MAE para la metodología propuesta y la convencional (Zhang et al., 2020)

En la comparación entre estas dos metodologías se obtiene un RMSE de 0.159 kWh y un MAPE de 12.68 %, lo que lleva a una mejora del 32.05 % para el RMSe y 30.14 % para la metodología planteada en el paper.

Capítulo 4

Metodología

4.1. Caso de Estudio

En esta sección se presentará la metodología que se siguió en el presente trabajo, el cual lleva como principales etapas, la creación de un modelo de consumo energético por kilometro, creación de modelo de predicción de velocidad y el modelo integrado que toma la estimación del modelo de velocidad como entrada al modelo de consumo. El modelo de consumo utilizó data de 3 vehículos eléctricos y para la estimación de velocidad se utilizó la información proveniente de 383 automóviles conducidos durante un año en Ann Arbor, Michigan, USA.

Para este estudio, los EV corresponden a Nissan Leaf del año 2013 con 24kWh de autonomía (Ficha Técnica Anexo 1). La base de datos esta conformado por cerca de 7500 kilómetros recorridos en la ciudad. Como se observa la figura 4.1 la altitud mínima y máxima son 220 y 330 metros sobre el nivel del mar, respectivamente, esto indica que los efectos de la inclinación no están bien representadas en la data.

Las temperaturas de la zona rondan entre los -15°C / 10°C en invierno y 15°C / 30°C en verano. Siendo la precipitación media de 3.2mm por día como se observa en las figuras 4.3 y 4.2. Esto muestra que el clima es de interior, medianamente seco, templado, pero presentando algunos días con mucha precipitación y días de nieve casi nulos. Esto da a entender el estado del pavimento debido a el clima no tiene un gran impacto en el estilo de conducción.

4.2. Base de Datos

La base de datos (DB) corresponde a una de gran tamaño, colectando data de 383 automóviles en Ann Arbor, Michigan US, donde, 264 vehículos son a gasolina, 92 HEV , 24 PHEV y 3 EV conducidos desde noviembre de 2017 hasta noviembre de 2018. Se capturan las trayectorias GPS de los vehículos, junto con sus datos de series de tiempo de combustible, energía, velocidad y energía utilizada por periféricos. En total se tienen cerca de 600 mil kilómetros de recorrido en distintas calles y diversos climas.

Como se observa en la figura 4.4, la data se compone de más de 32mil viajes, donde la media corresponde a viajes de 18.4 kilometro , donde cada uno está caracterizado por los siguientes parámetros, los cuales son comunes para todos los automóviles.

- VehID: Id de vehículo, las id para EV son 10, 455, 541

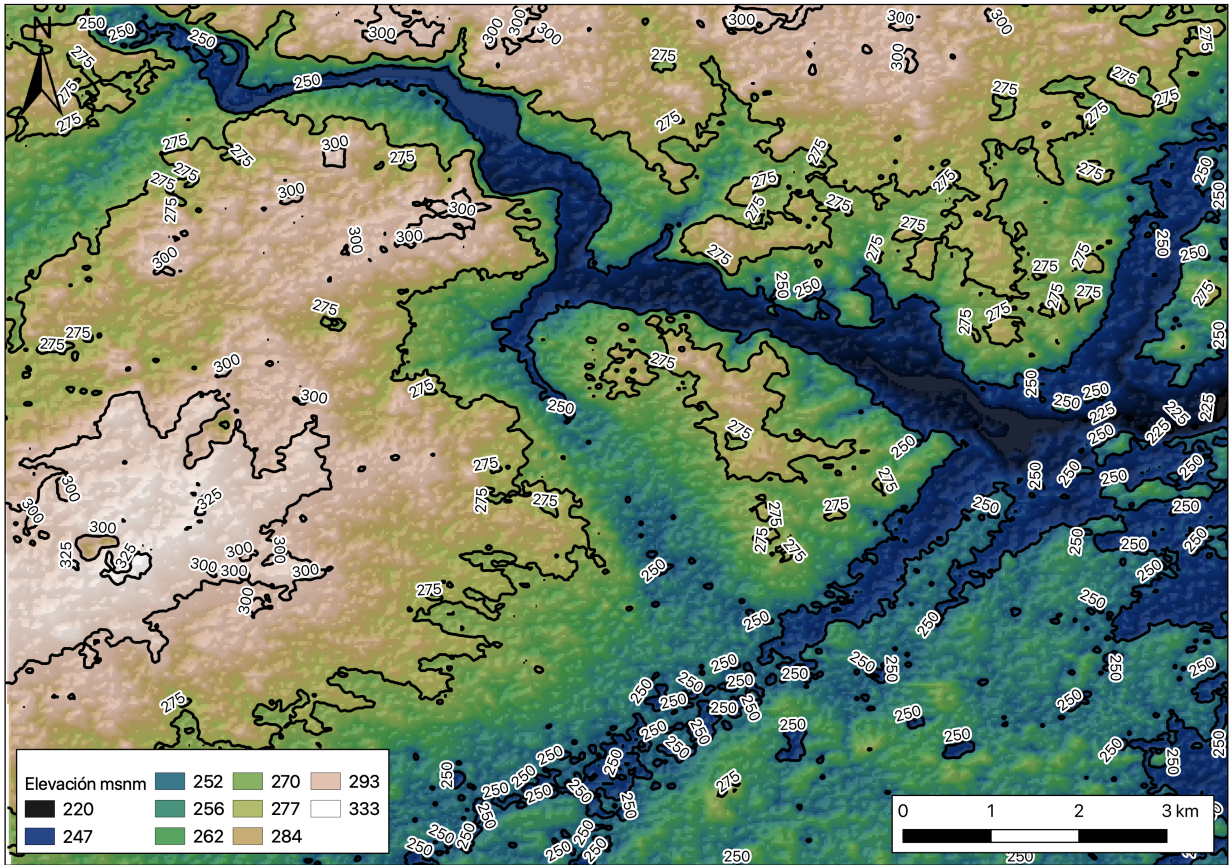


Figura 4.1: Mapa de elevación de la zona de Ann Arbor, Michigan, USA.

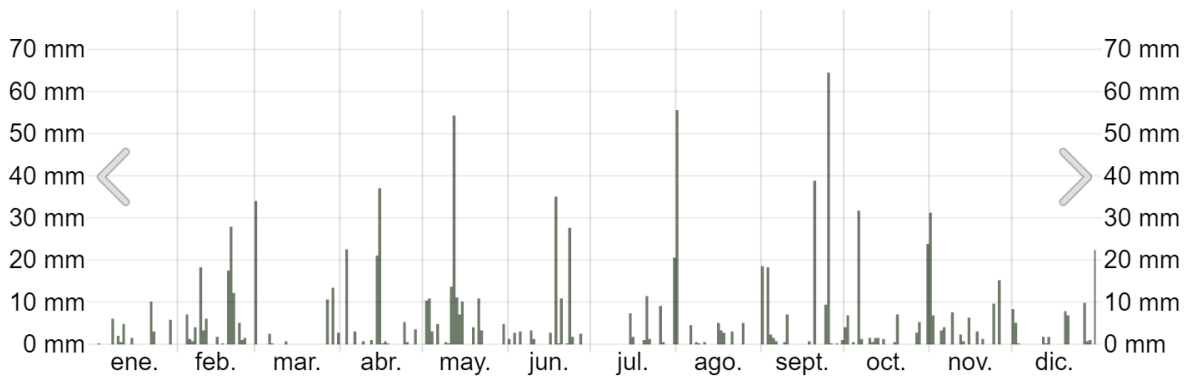


Figura 4.2: Precipitaciones año 2018 en Ann Arbor, Michigan, USA (Weatherspark, 2022.).

- Trip : Id del viaje realizado
- Latitude[deg]/Longitude[deg]: Coordenadas geográficas
- Timestamp(ms): Tiempo del viaje
- Vehicle Speed[km/h] : Velocidad del vehículo
- Air Conditioning Power[Watts]: Energía consumida por el aire acondicionado.

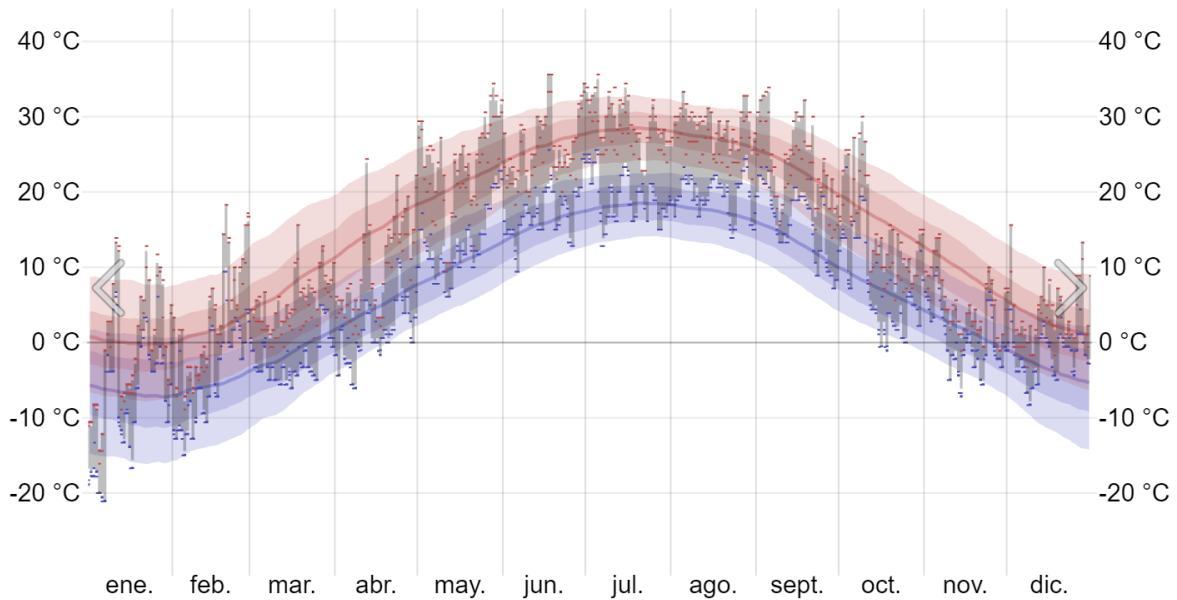


Figura 4.3: Temperaturas año 2018 en Ann Arbor, Michigan, USA (Weatherspark, 2022)

- Heater Power[Watts]: Energía consumida por la calefacción al interior del auto.
- Hora: Se entrega la hora en formato decimal

Los parámetros propios de los EV son los siguientes:

- HV Battery Current[A]: Corriente en batería
- HV Battery SOC[%]: Estado de carga de batería
- HV Battery Voltage[V]: Voltaje en batería

4.3. Extracción de Características

Basado en la fenomenología del modelo de consumo de un EV presentado en la sección anterior, además de los parámetros del diagrama de Ishikawa presentados en la figura 3.2, se realiza la extracción de características general que permite el posterior entrenamiento de los modelos, como se presenta en la figura 4.5.

- Hora del día: Se realiza la transformación numero-decimal a fecha y hora.
- Ajuste de ruta a camino más cercano: Debido a el error asociado a GPS que obtiene la latitud y longitud, esta se ajusta a la calle más cerca utilizando la API HERE.
- Clase de calle transitada: Luego de ajustar latitud y longitud, se obtiene el nombre y la clase de calle transitada utilizando la API OpenStreetMap.
- Submuestreo cada 3 seg: La data tiene 5 muestras cada 3 segundo, por lo que se submuestra para tener solo una muestra cada 3 segundos, utilizando la media como estimador.

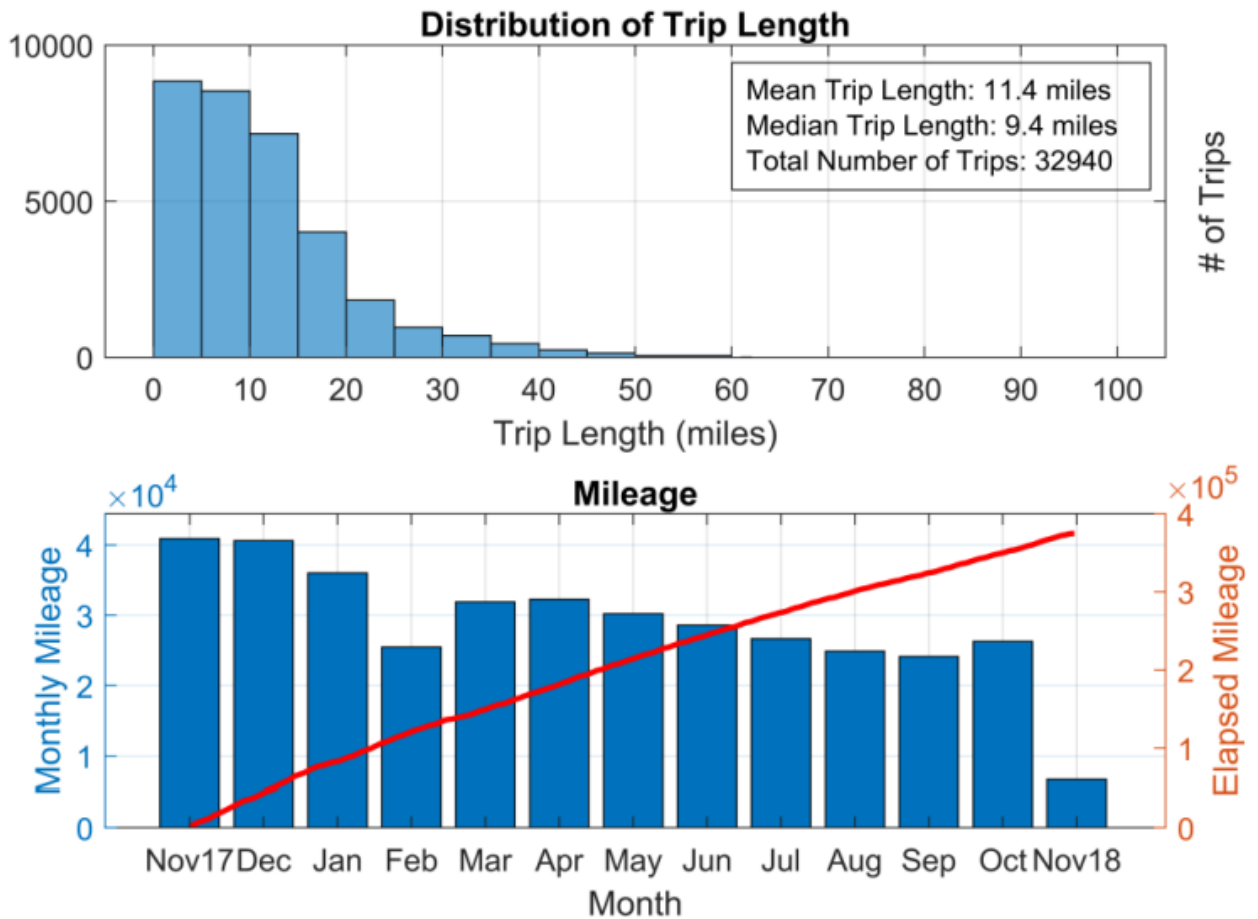


Figura 4.4: El gráfico superior muestra la distribución del largo de los viajes y la figura inferior la distribución de el total recorrido por mes(Oh et al., 2019)

- Extracción de nodos de interés: Utilizando la API OpenStreetMap se obtiene las coordenadas geográficas de los nodos de tránsito que se encuentran en la ruta. Se toma como nodos, los discos pare, ceda el paso y cruce peatonal (Se descarta el semáforo debido a que es una variable aleatoria que no se puede estimar)
- Elevación: Utilizando la API OpenStreetMap se obtiene la elevación de la calle transitada.
- Condiciones ambientales: Utilizando la API *Local Weather* de *World Weather Online*, se obtiene las condiciones ambientales en la localidad. Se toman las variables temperatura, humedad, precipitación, visibilidad, nieve.
- Distancia a nodo más cercano: Con la información anteriormente extraída, se obtiene la información de la distancia al nodo más cercano.
- Distancia a siguiente calle: Se realiza el mismo proceso que el punto anterior, pero con el cambio de calle en la ruta, la cual ayuda a estimar reducciones de velocidad.
- Variables Dinámicas: Se obtiene las variables dinámicas como $V(k-1)$, $V(k)$, $V(k+1)$, ángulo de elevación y energía consumida en un rango de tiempo.

Este proceso se realiza sobre la base de datos completa, a excepción de la obtención de la energía consumida, que se realiza sobre los datos relacionados con los EV. Estas variables se agrupan en una base de datos de variables general, que posteriormente se utilizan para modelos de estimación de velocidad y energía.

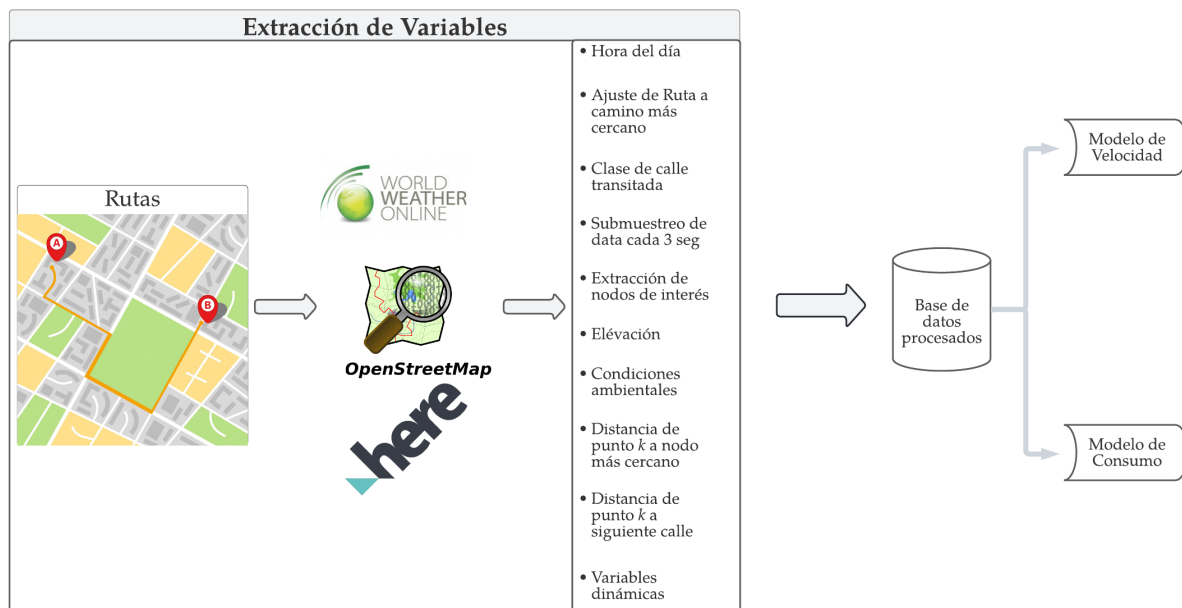


Figura 4.5: Diagrama de extracción de características general

4.3.1. Segmentación de Ruta

Con el fin de disminuir el ruido asociado a las mediciones de las variables dinámicas, es que se realiza un análisis por segmentación de ruta, esta no se realiza en un rango fijo, si no entre un rango que varía entre los 50 y 150 metros, ya que, en análisis posteriores se establece que es el rango donde se minimiza el error del modelo de consumo, como se observa en la tabla 4.1.

Para establecer este rango, se divide la ruta en tramos de 100 metros, pero si la distancia de un tramo es menor a 50 metros se une al tramo anterior, para evitar generar estadísticos con pocos datos. Es por esta razón que en la figura 4.6 se observan que se pasa de 4 tramos a 3, donde el primer tramo es de 100 metros el segundo es de 120 metros y el último es de 50 metros.

4.4. Modelo de Estimación de Velocidad

La propuesta para la solución del problema se divide en tres etapas, una de estimación de velocidad basado en data histórica e información del camino, otra de estimación de consumo utilizando data histórica de EV y la última que integra la información del primer modelo en el segundo.

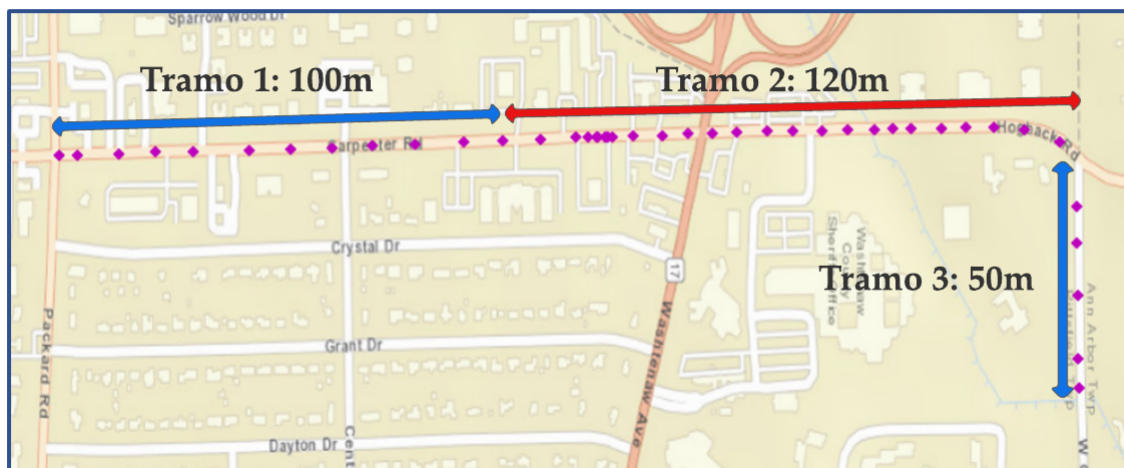
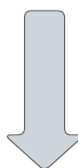
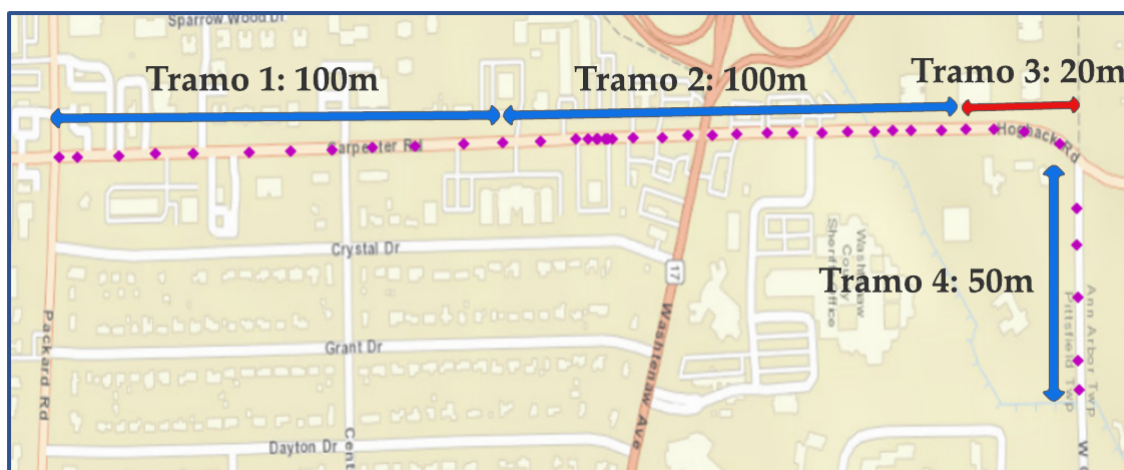


Figura 4.6: Ejemplo de segmentación de ruta en un rango $\in [50$ y $150[$ metros.

Tabla 4.1: Tabla de análisis de rango utilizando modelo de consumo

Rango [m]	RMSE [kWh/km]	MAE[kWh/km]
[0,100[0.0587	0.0544
[25,125[0.0387	0.0347
[50,150[0.0286	0.0274
[75,175[0.0290	0.0281
[100,200[0.0293	0.0279

En esta sección se describe el proceso para construir el modelo que estima la velocidad dado una ruta. Para este modelo se utiliza la data sub-muestreada cada 3 segundos de toda la base de datos, generando tramos como se especifica en la sección 4.3.1, incluyendo todos los tipos de vehículos, ya que se considera que el tipo de vehículo no condiciona la velocidad de conducción.

4.4.1. Estructura de características de modelo

Con el fin de establecer las características utilizadas, a modo de ejemplo, se presenta la ruta de la figura 4.7. donde se definen los siguiente parámetros:

- Indicador Rojo: Punto de partida
- Indicador Verde: Punto donde finaliza el viaje
- Indicador Azul: puntos de muestras en ruta, donde se extraen las características necesarias para la creación de los modelos.
- Nodo: Corresponde a un posible evento en ruta que puede cambiar la velocidad del EV, en el caso de la figura 4.7, este nodo corresponde a un disco pare.
- Nombre de calle: Corresponde a un identificador de la calle por la que se transita, la cual se utiliza para extraer indicadores como distancia a siguiente calle y tipo de calle transitada (primaria, secundaria, terciaria, residencial).
- Condiciones climáticas: Se toman las variables de Humedad, precipitación, visibilidad y nieve.
- Inclinación: Se toma la inclinación entre en punto k y el $k+1$.

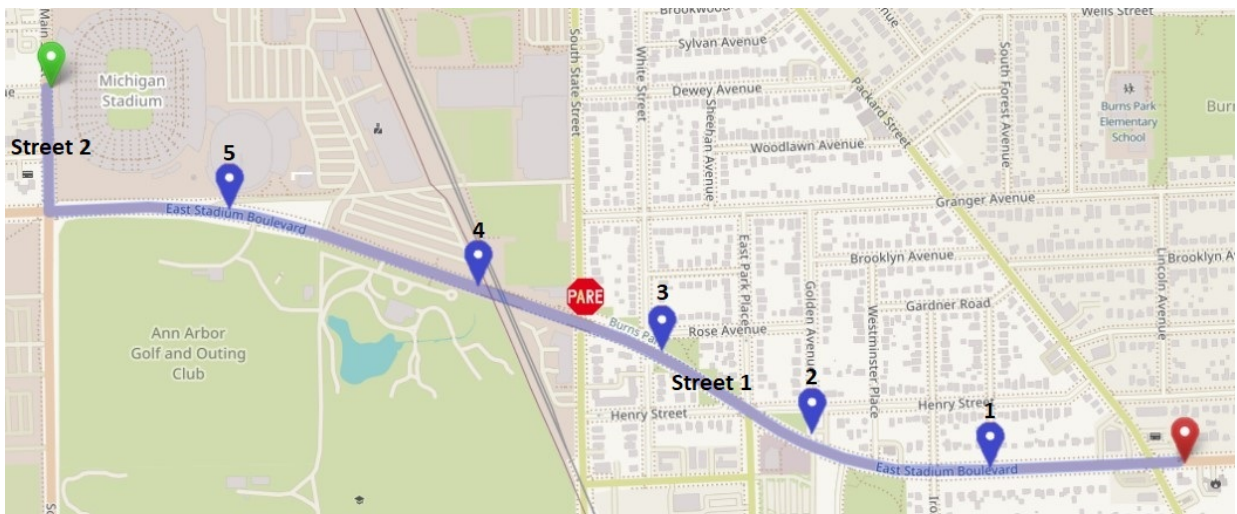


Figura 4.7: Ruta de ejemplo

4.4.2. Creación de Modelo

Una vez establecidas las variables que caracterizan la velocidad en ruta se procede a entrenar un modelo utilizando una red LSTM, cuya estructura se presenta en el Anexo 3, donde se realiza el estudio del tamaño de la ventana de variables para estimar $V(k)$. Como se observa en el Anexo 2, las características utilizadas son intrínsecas a la ruta, esto permite tener un ventana que asocie las características tanto en pasos anteriores como en pasos futuros, sin afectar la causalidad del problema.

En la tabla 4.2 se observa que el tamaño óptimo es de 2 pasos adelante y atrás, ya que, es donde hay una reducción de error considerable en comparación a la ventana de 1 paso.

Tabla 4.2: Error de red LSTM utilizando distintos regresores

Regresores	RMSE	MAE
k-1,k,k+1	20.1	20.4
k-2,k-1,k+1,k+2	13.1	13.9
k-3,k-2,k-1,k,k+1,k+2,k+3	12.9	13.4

Dado el análisis anterior, se define que $V(k)$ mediante la ecuación 4.1, cuyo proceso completo se ejemplifica en el diagrama de la figura 4.8.

$$V(k) = f_{LSTM}([data(k-2), data(k-1), data(k), data(k+1), data(k+2)]) \quad (4.1)$$

Como el proceso de estimación de velocidad resulta complejo de estudiar debido a la cantidad de variables aleatorias que se tiene en cuenta, es necesario caracterizar esta incertidumbre con el fin de obtener una mejor estimación de la salida. Como se establece en la sección 2.12, una forma de caracterizar esta incertidumbre es mediante el proceso *Monte Carlo Dropout*, obteniendo una distribución de velocidad para cada punto como se observa en la figura 4.8.

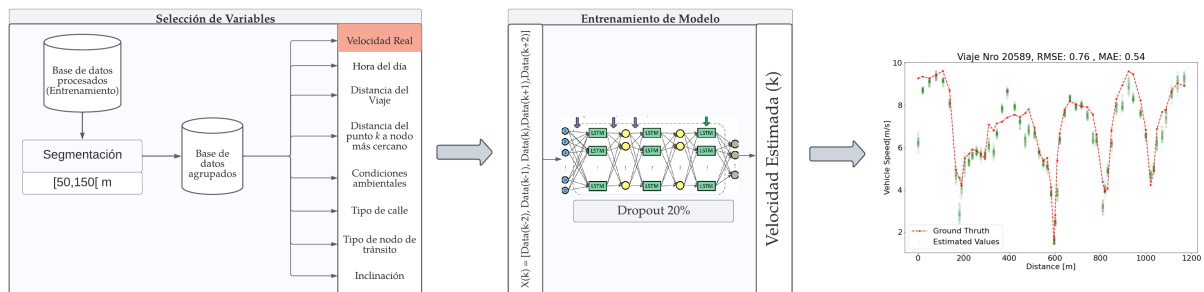


Figura 4.8: Diagrama del proceso de entrenamiento del modelo de velocidad utilizando una red LSTM. El parámetro remarcado indica el target que se quiere estimar.

4.5. Modelo de Consumo de Energía

La segunda parte del sistema consiste en la creación de un modelo de consumo basado en data, para esto primero se realiza una investigación de la fenomenología del problema a resolver, de esta forma se obtiene información sobre las posibles características que ayuden a representar el consumo en los futuros modelos a entrenar. Luego se procede a la extracción de estas características, principalmente de la ruta, donde a partir de las coordenadas y utilizando distintas API se consigue extraer todas las características correspondiente al estudio fenomenológico. A partir de este estudio se procede a entrenar un modelo basado en aprendizaje de máquina, es específico, árboles de regresión.

4.5.1. Estructura de Características del Modelo

Como se observa en la sección 3.2, existen diversas variables que impactan al consumo de un EV, entre estos se divide el consumo por tecnología, conducción, error en los instrumentos de medición, tipo de viaje, ambiente artificial y natural, como se especifican más en detalle en la figura 3.2. Teniendo esto en cuenta y tomando la ecuación 3.1 que analiza el consumo desde el punto de vista físico de un modelo simplificado, en este se observa como la velocidad y los cambios de velocidad (aceleración) tiene un gran impacto en el consumo. Basado en esto, el modelo utiliza las siguientes variables:

- **Velocidad** [m/s]($k-1, k, k+1$) : Se obtiene de la data real.
- **D(k+1)-D(k)**: Diferencia entre punto k y $k+1$, que hace referencia a la tasa de cambio de velocidad.
- **Inclinación** [°]: Obtenido de API *OpenStreetMap*.
- **Consumo de periféricos** [kW]: Se reemplaza por la humedad y temperatura, que tienen una directa correlación con el consumo de periféricos (Anexo 5)
- **Temperatura**: Obtenida de API *Local Weather* de *World Weather Online*

Los parámetros constantes como la densidad de aire, aceleración de gravedad, masa del vehículo, etc, se descartan por el hecho de que son combinaciones lineales de los componentes variables, por lo que no aportan en el entrenamiento basado en data.

4.5.2. Creación de Modelo

Una vez establecidas las características que representarán el consumo de energía en un EV, se procede a formar la base de datos con el que se entrenará. Con el fin de aumentar la información disponible y caracterizar de forma más completa el espacio de decisión, es que se realiza un proceso de *data augmentation* donde se completa la información agrupando la base de datos procesada en distintos rangos, como se observa en la figura 4.9, de esta forma se logra aumentar el espacio representado pasando de 20 mil filas a más de 100 mil para el entrenamiento.

4.6. Modelo Integrado

En las secciones anteriores se establece como fueron entrenados los modelos por separado. En esta etapa se realiza el proceso de validación, donde las variables dinámicas asociado a la

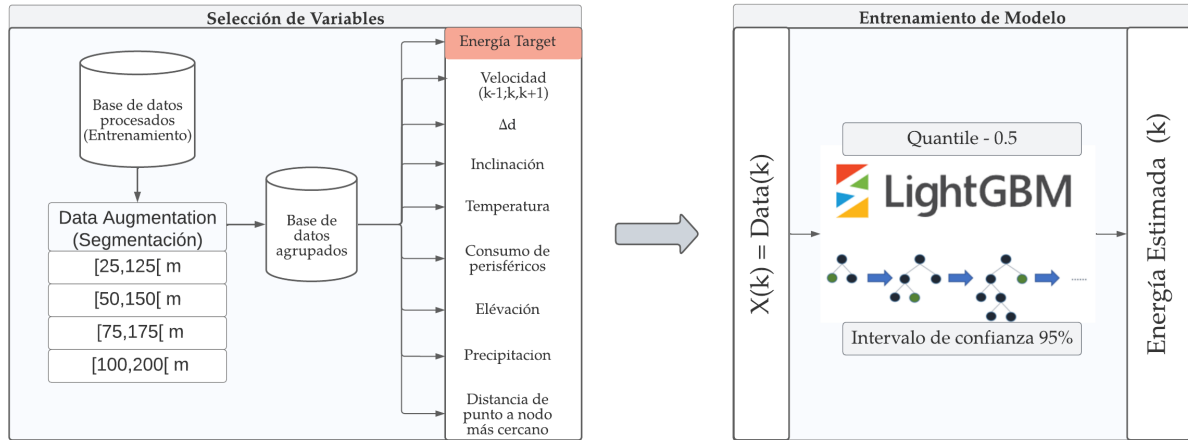


Figura 4.9: Diagrama del proceso de entrenamiento del modelo de consumo utilizando un árbol de regresión *LightGBM*. El parámetro remarcado indica la variable a estimar

velocidad del EV son estimados por el primer modelo realizando un análisis de incertidumbre utilizando *Monte Carlo Dropout*. A la salida de este primer modelo se obtiene una distribución punto a punto, luego estas variables se utiliza para la validar el modelo de consumo. De esta forma se entrega toda al incertidumbre a la variable dinámica más importante, obteniendo una distribuciones de consumo de energía por rango, para cada punto de las rutas de validación analizadas.

Según lo que se describe en la sección 4.5, el modelo de consumo se entrena utilizando la función de pérdida *quantile* (q), esto lleva a que se tienen que validar 3 modelos, para el rango del 50 % y el intervalo de confianza del 95 % ($q-0.025$, $q-0.5$, $q-0.975$, respectivamente) como se muestra en la figura 4.10. Es así como todos los modelos quedan bien representados, utilizando el valor medio de las distribuciones como estimador del valor del consumo energético por rango y su intervalo de confianza.

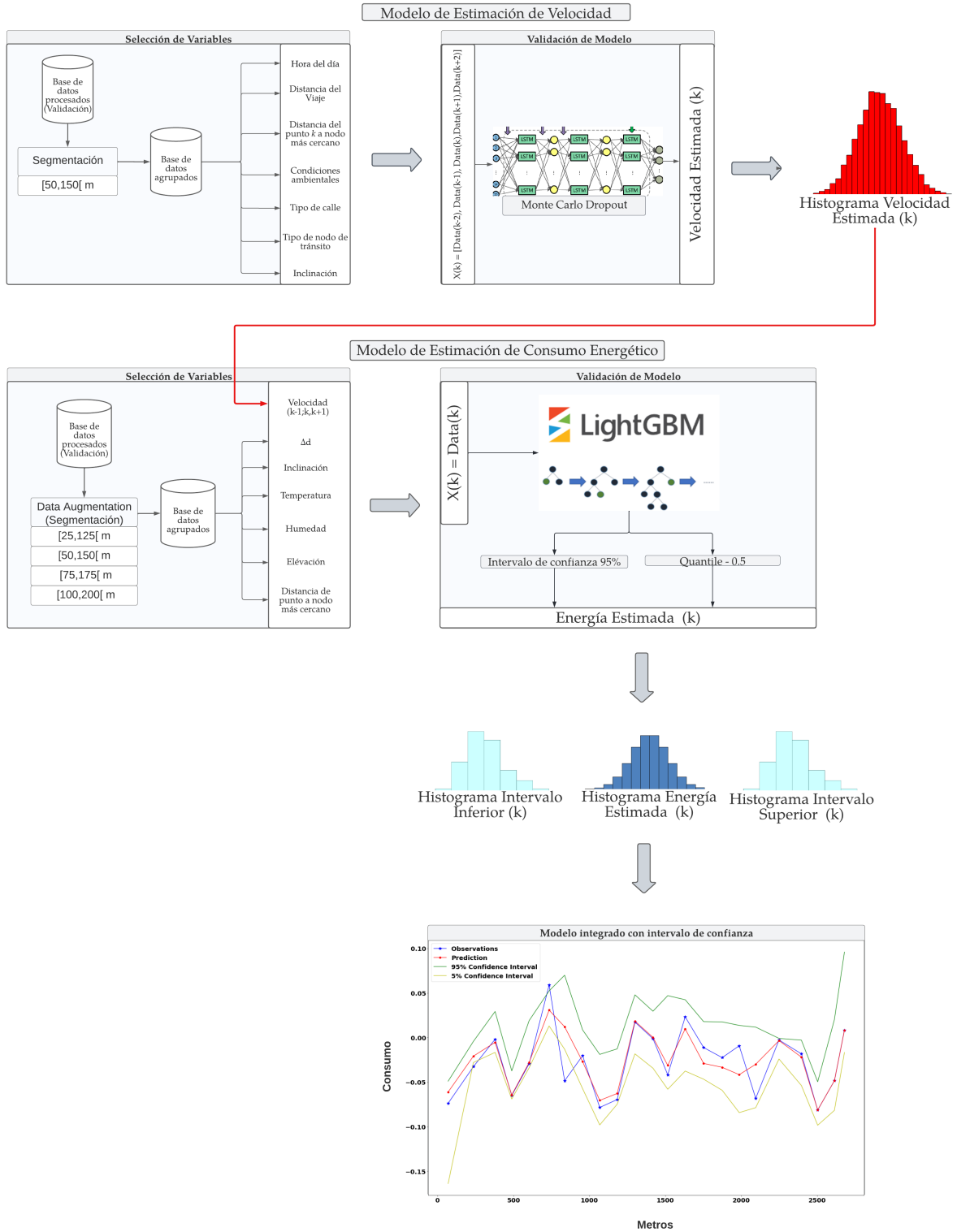


Figura 4.10: Diagrama del proceso de validación donde se unen el modelo de estimación de velocidad y el de consumo de energía.

Capítulo 5

Resultados y Discusiones

En este capítulo se presentarán y discutirán los resultados de la metodología planteada en el capítulo anterior. Este proceso se divide en 3 grandes secciones, el primero corresponde a los resultados del modelo de velocidad, donde se analizará los resultados de este modelo utilizando la librería SHAP que permite ver la sensibilidad del modelo a variaciones de las diferentes variables de entrada, entregando así un estimador del impacto de estas, así también su relevancia dentro del modelo. Por otro lado, en la segunda sección se presentará los resultados del modelo de consumo con variables de velocidad reales (extraídos de la base de datos), de esta forma se obtendrá el modelo base. En la última sección se mostrarán los resultados del modelo integrado, que conlleva una estimación de la distribución de velocidad, esto lleva a que la salida del modelo integrado entregue una distribución de consumo, donde se analizará el desempeño de la metodología completa y como se comporta para estimar tanto la energía consumida en cada rango de viaje, como la integral del consumo estimado total en la ruta.

5.1. Modelo de Velocidad

El modelo se validará sobre 300 viajes el cual corresponde a cerca del 1% de los viajes totales, donde se incluyen 34 EV y el resto son tomados de forma aleatoria, donde las métricas corresponden a la comparación entre el cuantil 0.5 de la distribución de cada punto y el valor real. Como se observa en la figura 5.1, este es cerca de un 7.38% lo cual corresponde a 1[m/s] o 3.6[km/h] en promedio, con esto queda en evidencia el buen desempeño del modelo de forma global.

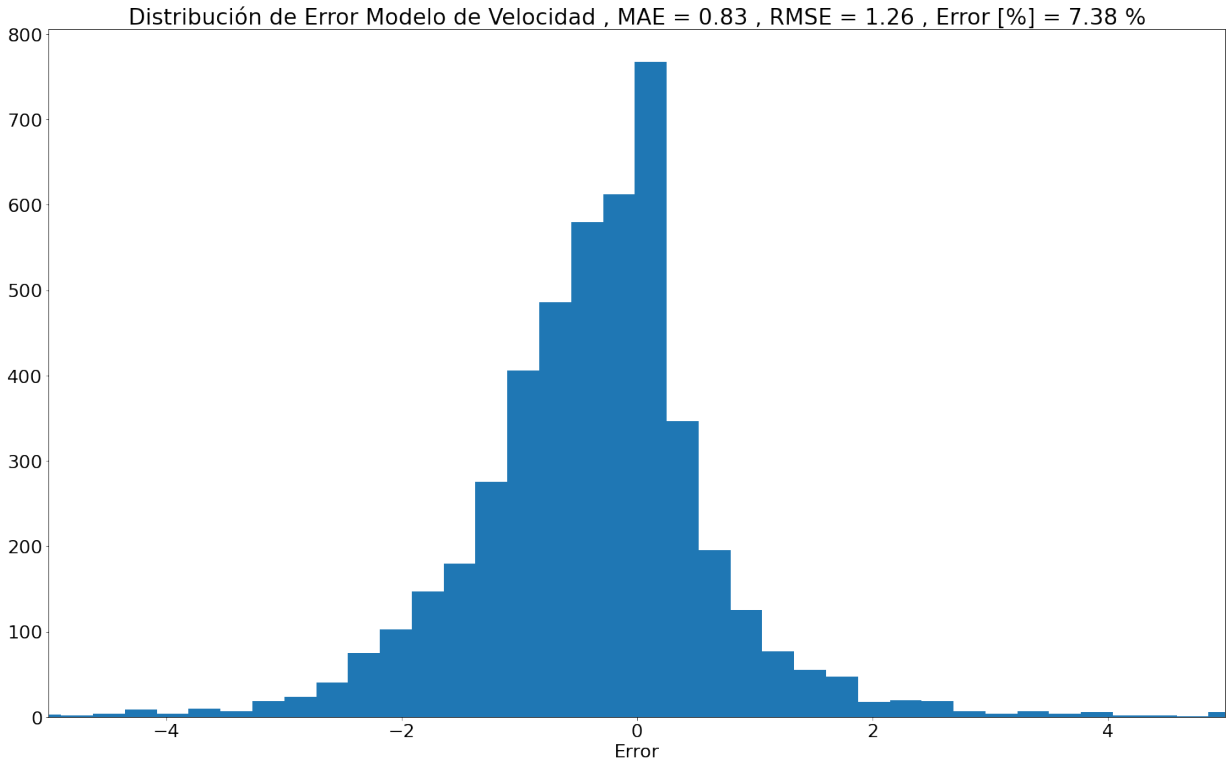


Figura 5.1: Error de modelo de velocidad sobre todo el conjunto de validación.

Para evaluar el desempeño del modelo de forma más específica, es que se planeta evaluarlo por rango de velocidades como se muestra en la figura 5.2, donde observando las distribuciones, los rangos sobre 80 [km/hrs] existe una menor cantidad de datos. Analizando más en profundidad la distribución de error en los distintos rangos, los resultados mostrados en la tabla 5.1, donde existe un mayor error es en las velocidades más bajas, esto se debe a que existe una mayor incertidumbre ya que corresponde a velocidades de partida o detención del vehículo, además de representar la velocidad más cercanas a zonas residenciales o terciarias, donde existen reducción de velocidad aleatorias asociadas a variables que no están consideradas en el modelo, como por ejemplo reductores de velocidad y semáforos . Por otro lado, observando la representatividad de la data en las distintas velocidades, es que se tiene que la data menor representada corresponde a las velocidades mayores a 80 [km/hrs] teniendo un error bajo en comparación a las velocidades más representadas, esto se debe a que estas son velocidades asociadas a calles principales, en especial autopistas, en donde existen menos variables aleatorias que influyan en la velocidad, como semáforos y reductores de velocidad.

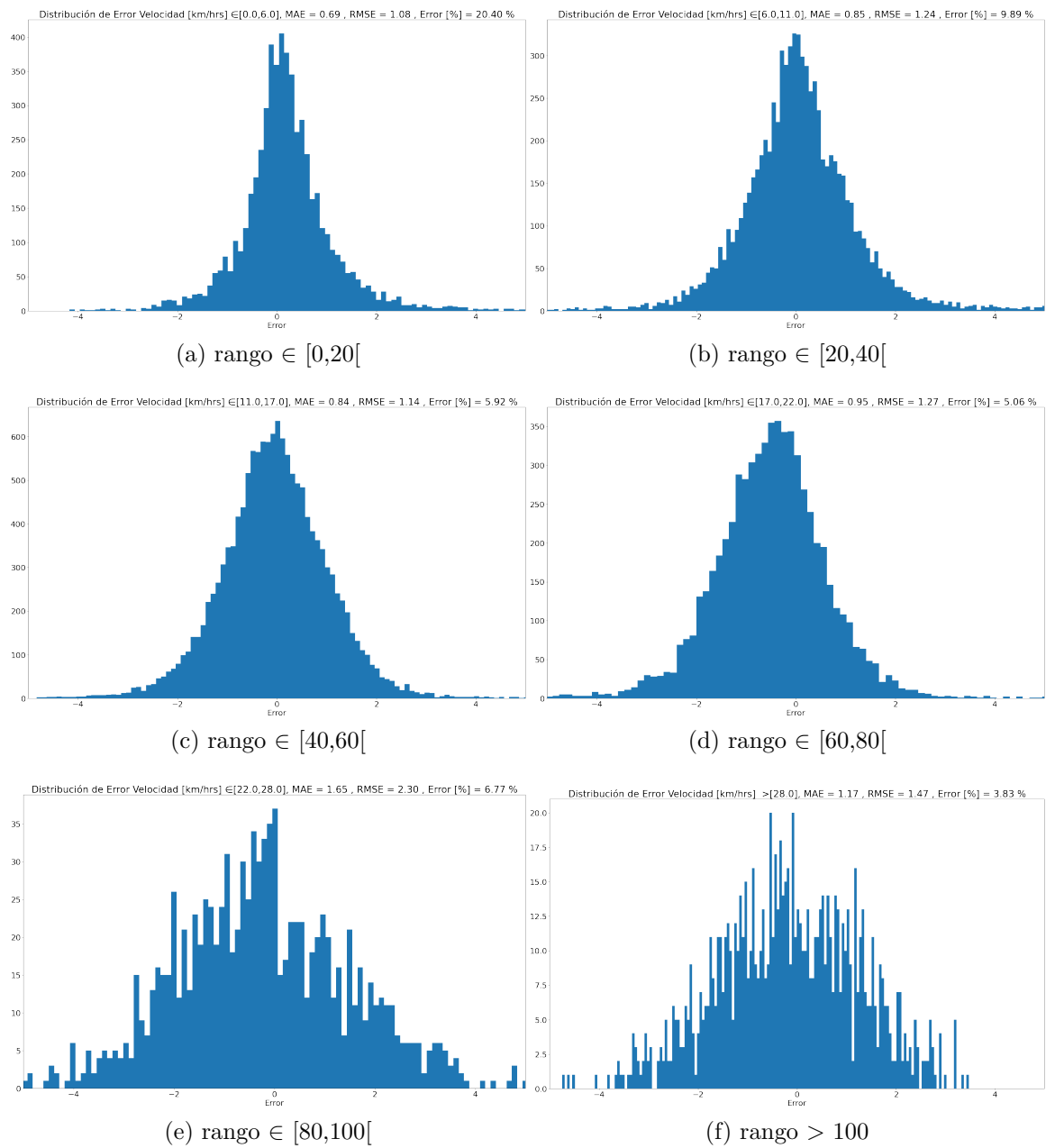
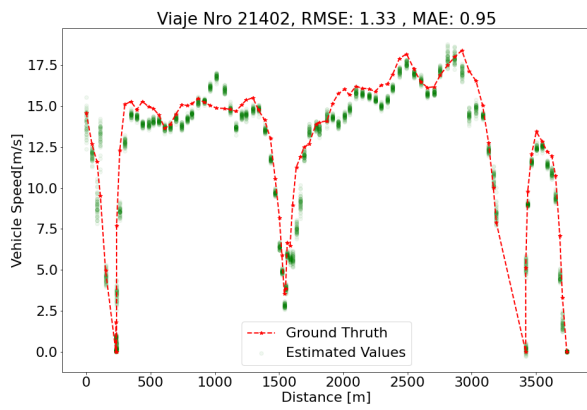


Figura 5.2: Distribución de error según rango

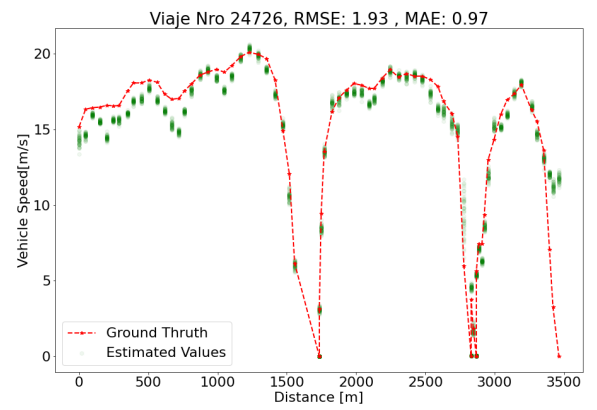
Tabla 5.1: Análisis de error por tramos de velocidad en conjunto de entrenamiento

Rango [m/s]	Rango [km/hrs]	RMSE [m/s]	MAE [m/s]	% Error	% Entrenamiento
[0,6[[0,20[0.68	1.07	20.40 %	14.47 %
[6,11[[20,40[0.85	1.23	9.89 %	19.37 %
[11,17[[40,60[0.83	1.13	5.92 %	38.10 %
[17,22[[60,80[0.95	1.27	5.05 %	16.48 %
[22,28[[80,100[1.65	2.29	6.76 %	2.69 %
>28	>100	1.17	1.47	3.83 %	3.43 %

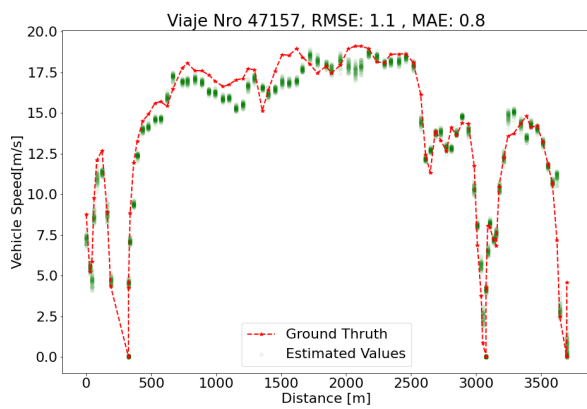
Analizando los viajes presentados en la figura 5.3, se observa como el error aumenta en los últimos tramos, es decir donde se pierde información de las variables asociadas al suavizado ($k+1$ y $k+2$), debido a esto es que el viaje (b) presenta un error cercano 10 [m/s] o 36[km/hr]. Por otro lado las estimaciones al inicio del viaje son precisas en la media , pero la distribución de la velocidad aumenta debido a la falta de información de las variables asociadas al filtrado.



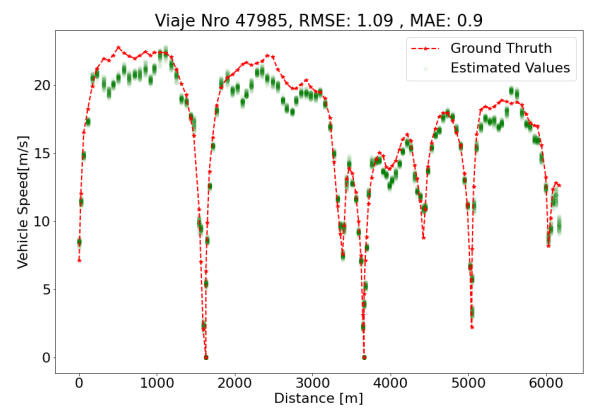
(a)



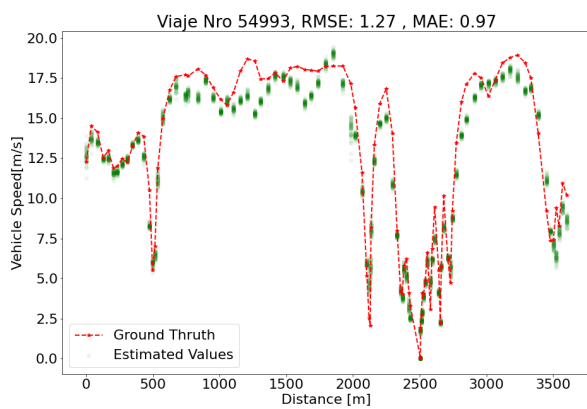
(b)



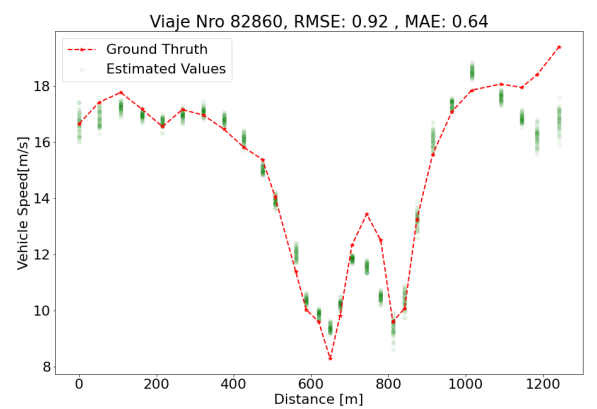
(c)



(d)



(e)



(f)

Figura 5.3: Distribución de error según rango

Con el fin de comprender el impacto de cada variable, es que se presenta el análisis de la figura 5.4 correspondiente a los *SHAP values* de todas las variables, donde se aprecia como las variable de la distancia a la siguiente calle tiene un impacto muy grande, el cual aumenta

cuando la distancia es mayor, es decir la probabilidad de reducir la velocidad debido cambios de trayectoria disminuye.

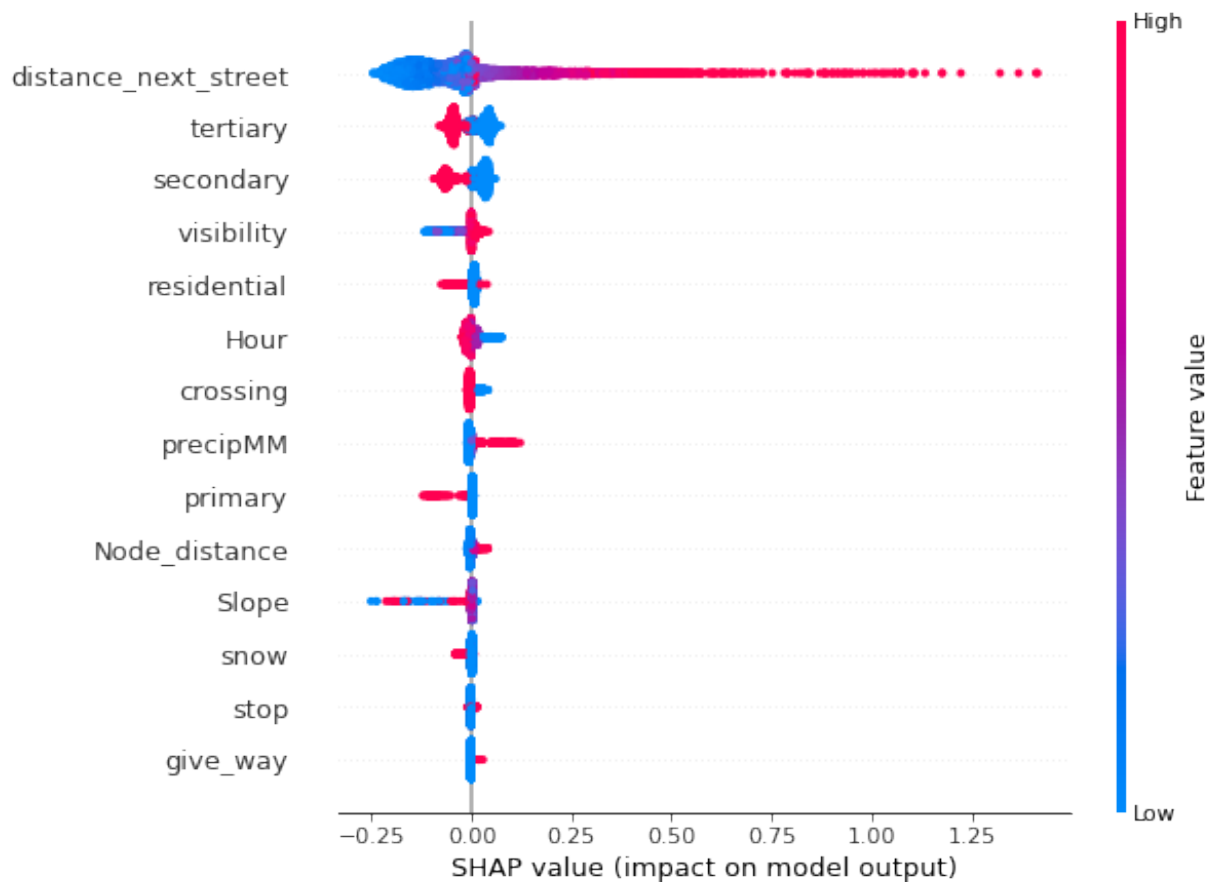


Figura 5.4: *SHAP values* de todas las variables sobre el modelo de velocidad.

Por otro lado, las variables asociadas a las condiciones climáticas y de horario, tienen un impacto menor pero no menos importante. Si la visibilidad aumenta tiene un impacto positivo en el aumento de la estimación, por otro lado, la nieve tiene el efecto esperado, donde a mayor valor de este, se tiene un impacto que disminuye el valor de la salida. En relación a hora se puede ver que a horas más tarde la velocidad disminuye y por otro lado a horas más temprano aumenta el impacto. En relación a las precipitaciones, esta tiene un comportamiento que va contra la heurística, que nos dice que a medida que aumenta la precipitación, esta tiene un impacto negativo, esto se puede deber a la poca data que tiene un impacto real en la zona, por lo que se puede decir que esta no genera un impacto real en el modelo.

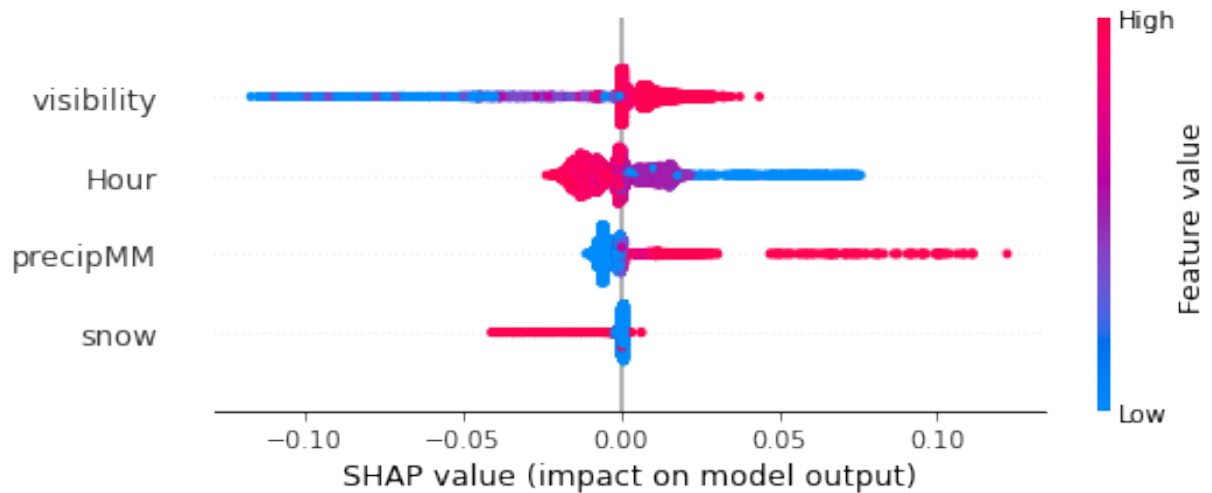


Figura 5.5: *SHAP values* de variables asociadas al clima sobre modelo de velocidad.

Analizando los *SHAP values* asociados a variables artificiales de la ruta, según el diagrama de Ishikawa de la figura 3.2. Las variables del tipo de calle, las cuales están codificadas utilizando *one hot encoding*, presentan un impacto negativo en la velocidad, esto se puede deber a que el cambio de una calle a otra está asociada con una gran disminución de velocidad, lo que lleva a que estas variables muestren un impacto muy negativo. Lo mismo sucede con las variables asociadas a la señales de tránsito, las cuales se codifican de la misma manera. Por otro lado una variable interesante es la distancia al nodo más cercano (*Node_distance*), donde a mayor distancia del nodo más cercano aumenta la velocidad estimada, lo cual tiene relación con la fenomenología del problema.

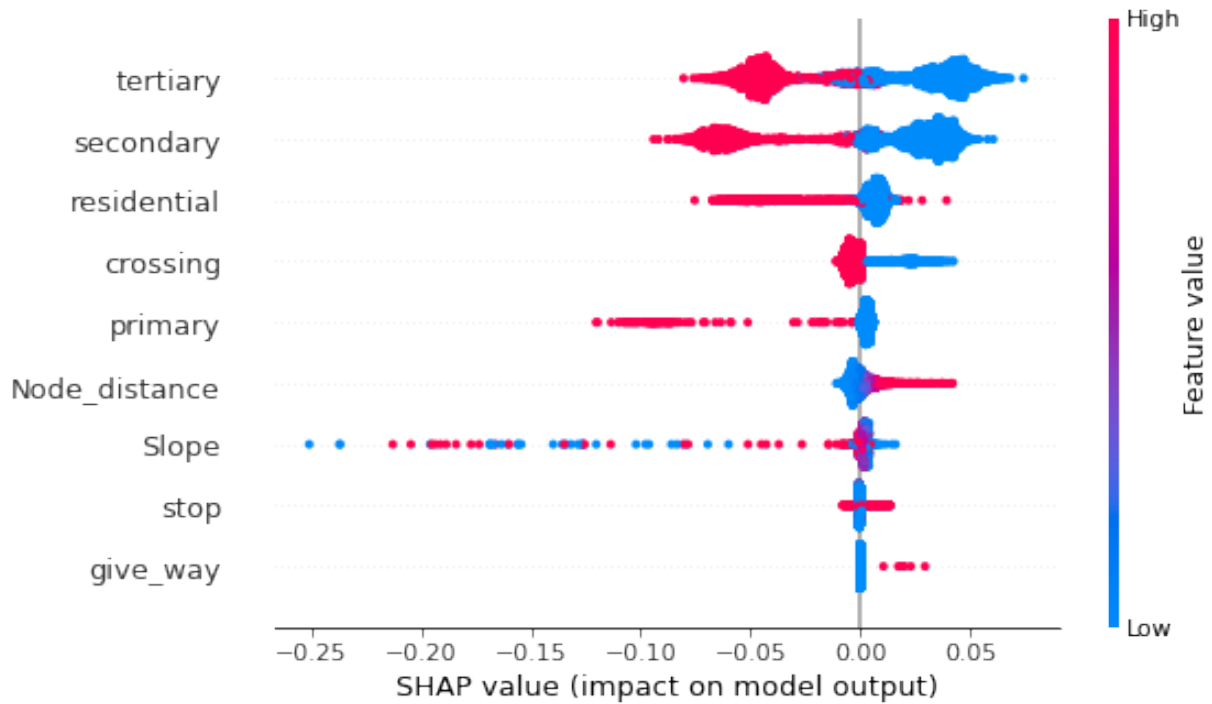


Figura 5.6: *SHAP values* de variables artificiales de la ruta.

5.2. Modelo de Consumo

El análisis a continuación se realiza sobre 34 viajes para la validación, que corresponde a 1.47% de los viajes totales del dataset. Este análisis se realiza con la data real de variables dinámicas, es por esto que este sería el caso base, es decir, el mejor resultado que se puede conseguir si el modelo de estimación de velocidad hiciera una estimación perfecta.

Observando los viajes presentados en la figura 5.7, donde de manera general el modelo tiene buen desempeño estimando el consumo, pero el error aumenta al estimar los efectos del freno regenerativo. Esto se debe a la complejidad del consumo, donde resulta difícil integrar todas las variables asociadas a este (Zhang et al., 2020), debido a la cantidad de consideraciones que se afectan con distinta intensidad, donde según el diagrama de Ishikawa (figura 3.2) hay múltiples variables que no se tienen en consideración debido a las limitantes de la DB, como variables asociadas al conductor, error en mediciones, entre otras.

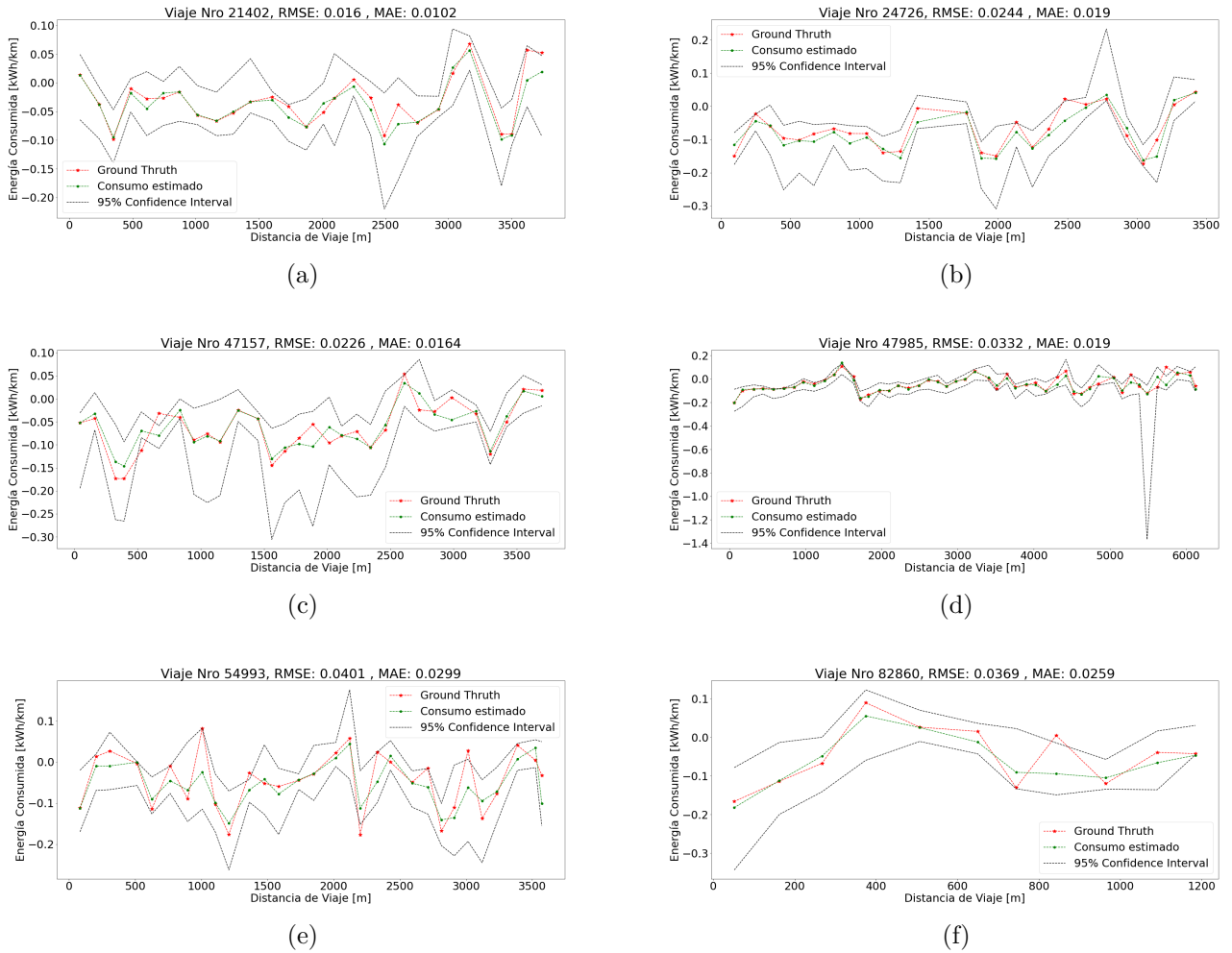


Figura 5.7: Distribución de error según rango

Con el fin de tener una comprensión más profunda del desempeño del modelo, es que se analizan las distribución del error global, además de separarlo por consumo y regeneración, como se muestra en la figura 5.9 y cuyo resumen se observa en la tabla 5.2 . En cuanto al error global, se tiene un 27% en la estimación por tramo y si se analiza por consumo y regeneración, queda en evidencia el análisis de los viajes de la figura 5.7, donde el error para el freno regenerativo (figura 5.9 (a)) es casi 3 veces mayor que el error asociado al consumo (figura 5.9 (b)). Esto en parte se debe, a que el la cantidad de data disponible para consumo es casi 3 veces mayor, por lo que el los efectos del freno regenerativo está menos representado que el consumo. Por otro lado, como se mencionó anteriormente, no se tiene en consideración efecto el tipo de conductor que regula el tipo de conducción el cual tiene un gran impacto en el consumo (Felipe et al., 2015), además de otras variables asociadas al desgaste del EV.

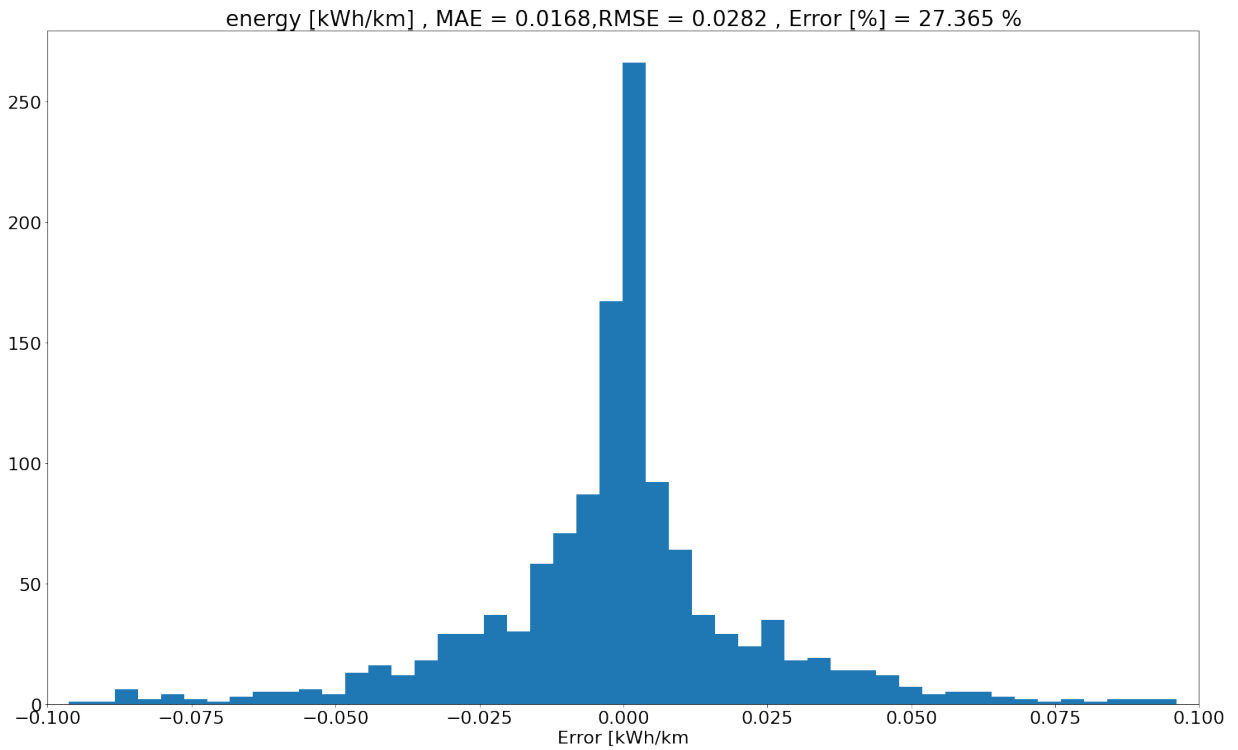


Figura 5.8: Error de modelo de consumo sobre todo el conjunto de validación (caso base), utilizando variable de velocidad real.

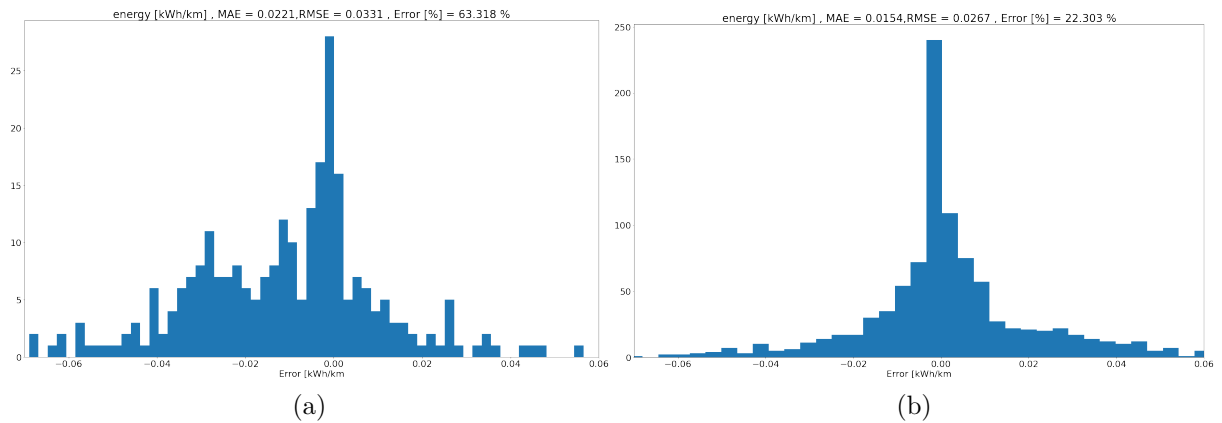


Figura 5.9: (a) Distribución de error para freno regenerativo debido a los efectos del freno regenerativo (b) Consumo negativo debido a los cambios de velocidad positivos

Tabla 5.2: Análisis de error por tramos de consumo

Rango	RMSE [kWh/km]	MAE [m/s]	% Error	% Entrenamiento
>0	0.0331	0.0221	63.3 %	23 %
<=0	0.0267	0.0154	22.3 %	77 %

Con el fin de comprender como impactan las variables al modelo desarrollado, es que se estudian los *SHAP values* para tener una comprensión más profunda del modelo. En este caso mientras más negativo el valor, es que tiene un mayor impacto en el consumo y si el valor es positivo es que tiene un impacto en la regeneración. Las variables asociadas a la dinámica del EV son la velocidad en $k-1, k$ y $k+1$, donde para valores alto de la velocidad en $k-1$ el impacto en al variable de salida es positivo, es decir, aparece el efecto del freno regenerativo y por lo contrario si la velocidad en $k+1$ es alto el impacto es negativo, es decir, aumenta el consumo, por otro lado si la velocidad en k es alta, el consumo consumo aumenta, pero no tiene un gran impacto como las velocidades en $k-1$ y $k+1$ las cuales hablan de los cambios de velocidad, componente que tiene un impacto un exponencial mayor a la velocidad por si sola, tambien hay que tener en cuanto la variable de diferencia de distancia (*Delta_d*), que da información si esta diferencia en $V(k-1)$ y $V(k)$ se produce en una distancia mayor o menor, donde a mayor sea la distancia el impacto es positivo. En relación a la inclinación, se aprecia que a mayor valor, mayor es el impacto, esto a pesar de que la zona no sea de alto relieve, demostrando que esta variable resulta muy importante en el consumo, lo cual se haría aún más evidente en zonas donde las diferencias topológicas sean mayores. Por otro lado se observa que el consumo de periféricos está muy asociado a la temperatura y humedad ($OAT[DegC]$).

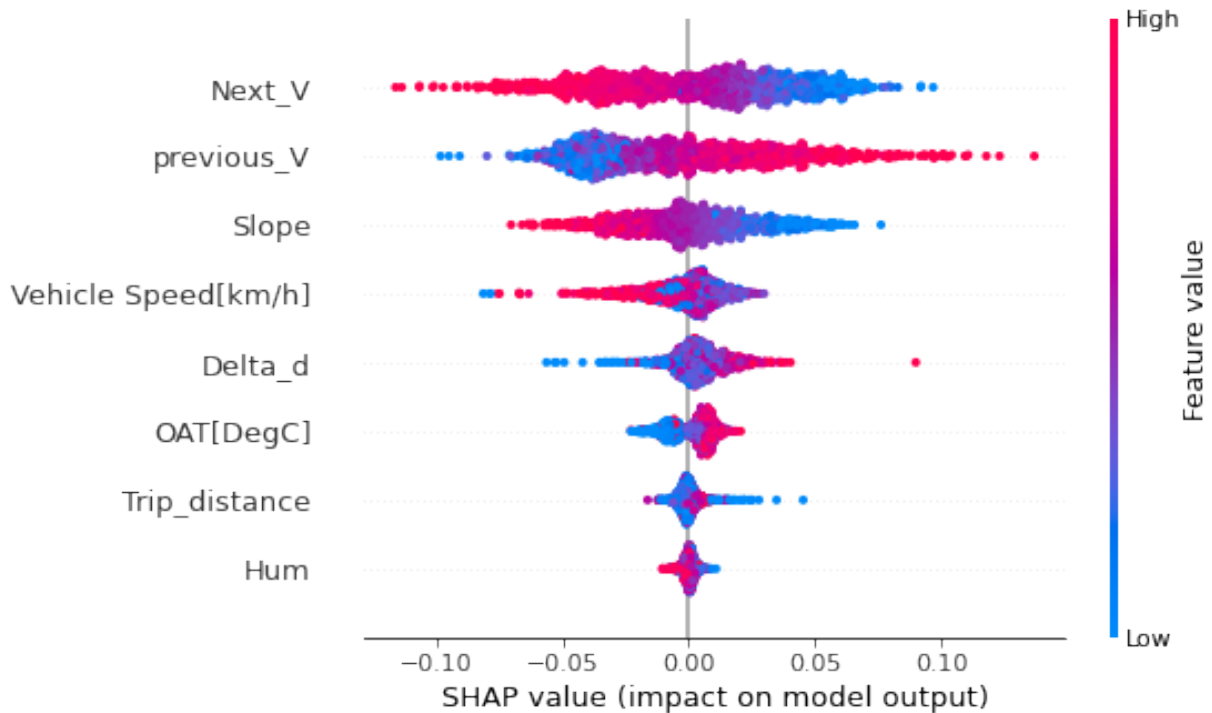


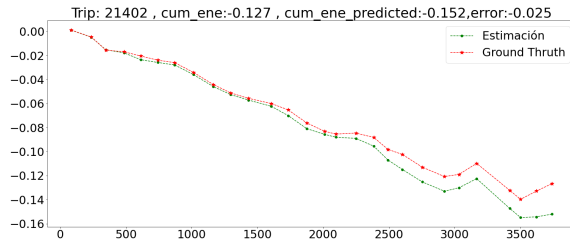
Figura 5.10: *SHAP values* modelo de consumo de energía

5.2.1. Análisis de Integral de Consumo

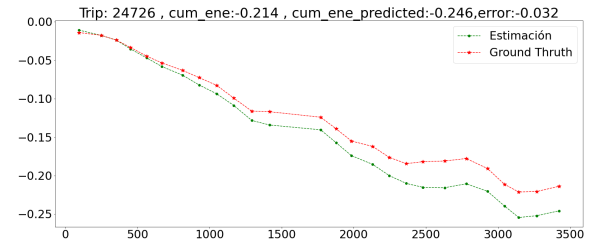
Utilizando el modelo que estima la energía consumida (kWh/km) en tramos de [50,150[metros, se realiza la integral sobre la distancia a cada punto estimado en ruta, de esta forma se obtiene el consumo acumulado, donde el último punto es la energía consumida total estimada. Analizando los gráficos que se muestra en al figura 5.11, se tiene que el consumo se ajusta de forma precisa al consumo real, por lo que si analiza el error asociado a la integral completa de todos los viajes de validación, se obtiene un 7.8% de error, lo que corresponde a un MAE de 0.0167.

Comparando los resultados de la integral del del consumo con la estimación punto a punto, es que se tiene un clara mejora en el error, llegando tener un error porcentual 3 veces menor, esto deja en evidencia que el error asociado al consumo punto a punto es mayormente gaussiano, haciendo que al integrarlo, se promedie, reduciéndolo significativamente. Por otro lado si toma el consumo final de la integral de las rutas de validación y se suman, se obtiene que el valor real es de -6.91 kWh y para el consumo estimado es de -7.11 kWh, teniendo un error de un 2.9%, el cual sobre estima el consumo real.

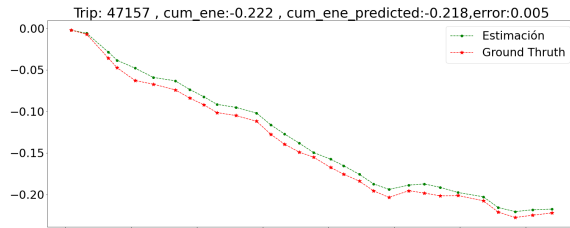
El análisis anterior nos lleva a que al integrar el modelo de consumo, el modelo integrado debería tener un buen desempeño siempre y cuando la estimación de velocidad tenga un error reducido.



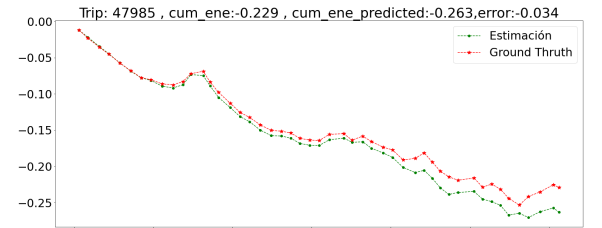
(a)



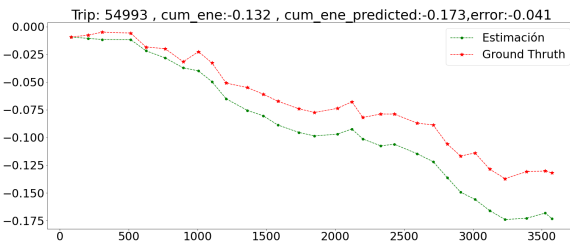
(b)



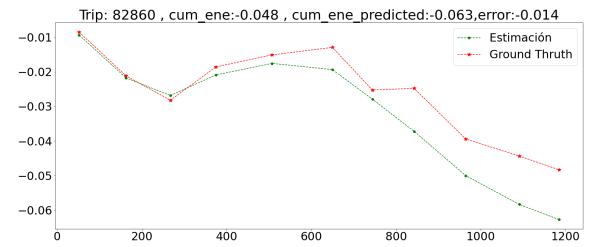
(c)



(d)



(e)



(f)

Figura 5.11: Integral del consumo, target vs estimación, para el modelo base con velocidad entregada por data

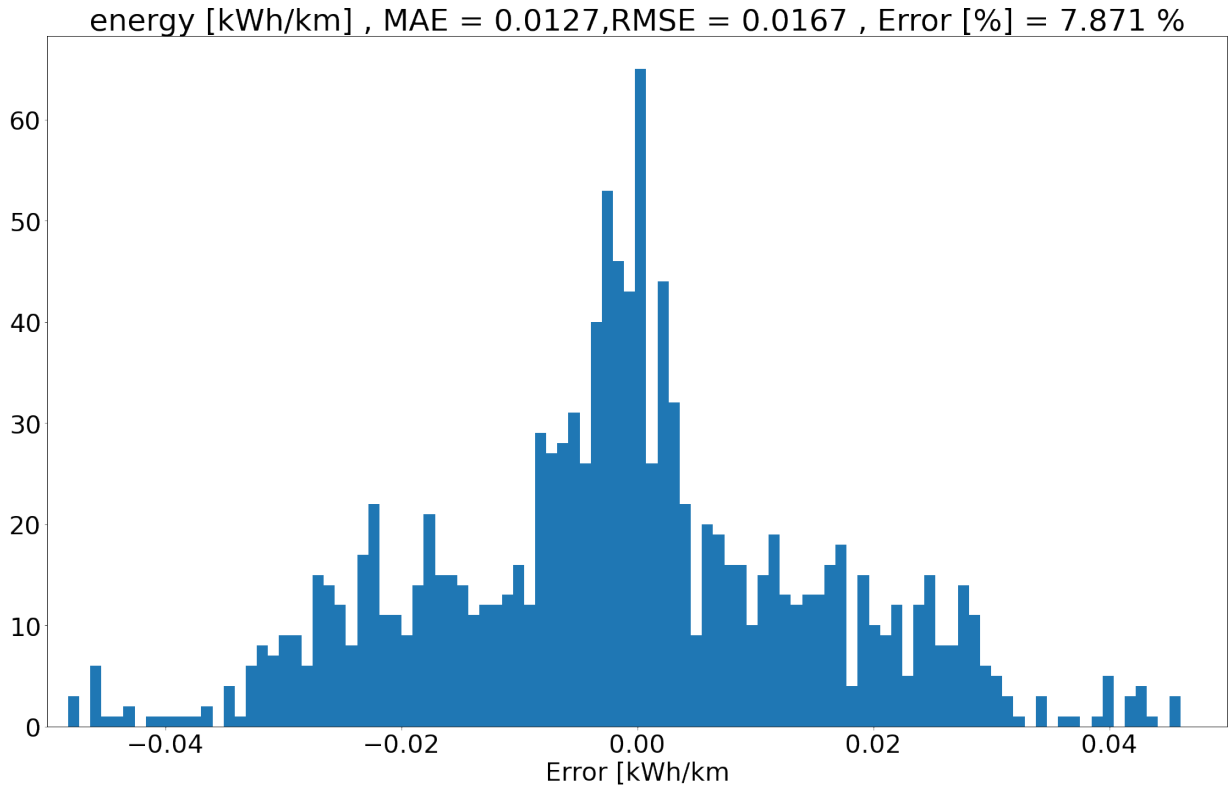
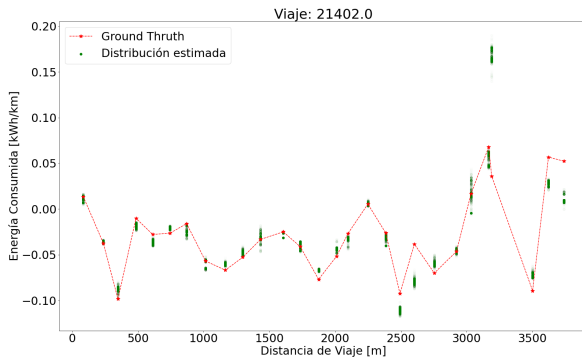


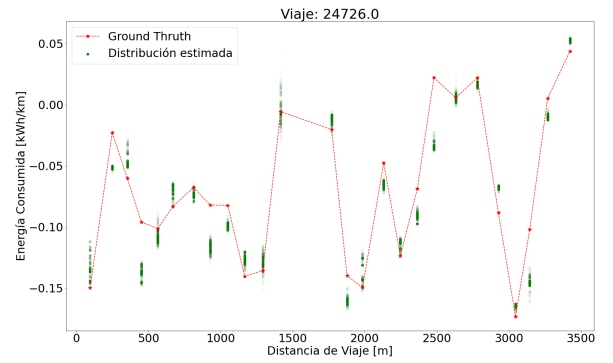
Figura 5.12: Error de modelo de consumo acumulado sobre todo el conjunto de validación (caso base), utilizando variable de velocidad real.

5.3. Modelo Integrado

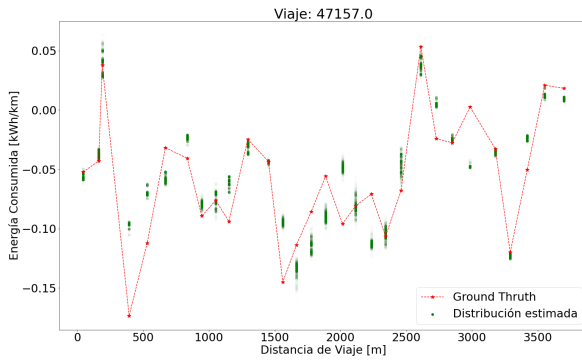
A continuación se presentarán los resultados del modelo integrado que se muestra en la figura 4.10, donde las variables dinámicas que son las que tienen más impacto son estimadas por el modelo de velocidad, entregando una distribución asociada a la incertidumbre del modelo. En la figura 5.13 se observan algunos resultados de la estimación del modelo completo, donde para cada punto se tiene una distribución asociada. En la mayoría de los casos esta distribución sigue una distribución gaussiana, pero en otros casos se observan acumulaciones en dos puntos, como en el último punto de la figura 5.13 (f), esto se puede deber a que existen variables como la distancia al nodo más cerca, siendo el nodo un paso peatonal, que son una variable aleatoria que puede o no tener un efecto en la reducción de velocidad, otro motivo es la falta de información de la data de consumo, donde no para toda la distribución de velocidad se tiene una salida mapeada, lo que provoca saltos de información en ciertas distribuciones.



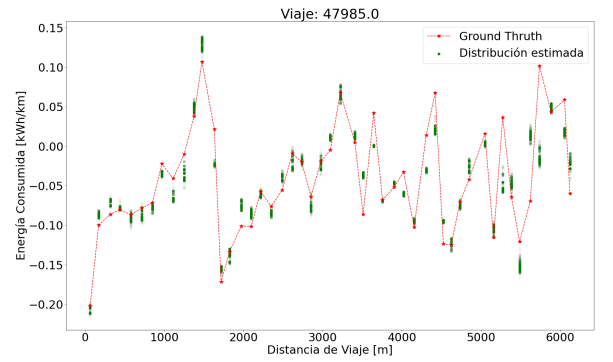
(a)



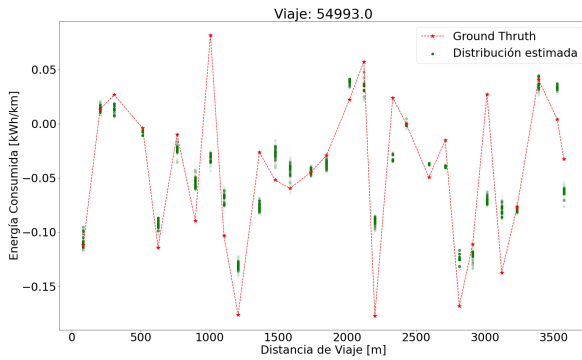
(b)



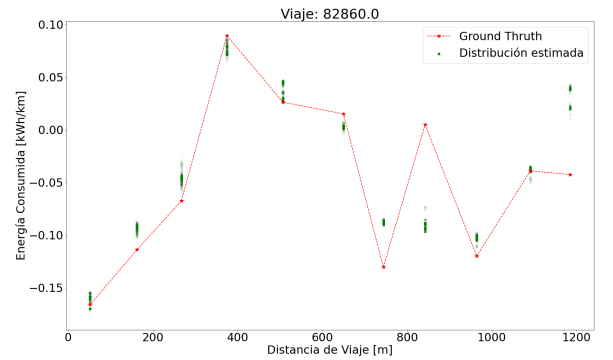
(c)



(d)



(e)

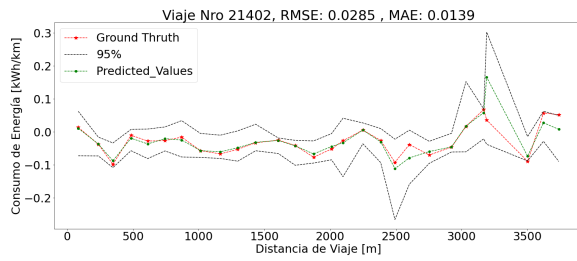


(f)

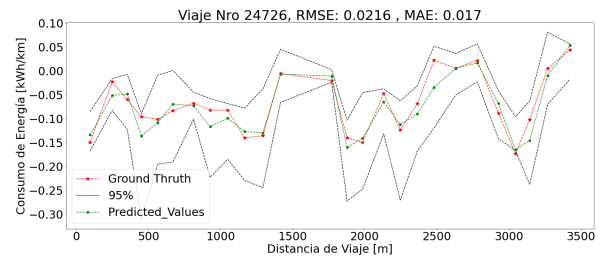
Figura 5.13: Distribución de modelo consumo instantáneo punto a punto, utilizando velocidad estimada a la salida dl modelo de velocidad.

Con el fin de comparar el desempeño con el caso base, es que se toma la media de la distribución para los intervalos de confianza y el valor asociado al quantile 0.5. Comparando los viajes se observa que en todos los casos el error aumenta de forma leve, para tener un punto de comparación estadísticamente correcto es que se compara la distribución asociada a todos los viajes de validación con el error del caso base. En la figura 5.15 se observa como hubo un aumento de error un 13 %, de acuerdo al valor medio de consumo, esto se traduce en un aumento de MAE 0.0077 [kWh/km] y un aumento de RMSE de 0.0069[kWh/km]. Esto quiere decir que para cada estimación de consumo por tramo se tiene que el modelo reduce su desempeño en comparación al caso base, lo cual es un resultado esperado, debido a la suma del error de ambos modelos.

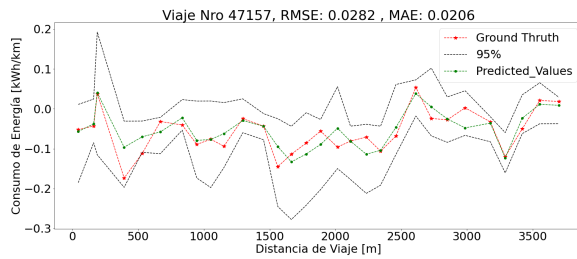
Con el fin de comprender que rangos de consumo se ven más afectados, es que se comparan las distribuciones de error de la figura 5.16 y el consumo por rango del caso base mostrado en la figura 5.9. En esta comparativa es que se observa un error de 81.5 %, lo que se traduce en un aumento del 18.2 % en relación al valor medio del efecto del freno regenerativo, en cuanto al error asociado al consumo, es de 34.1 % lo que se traduce en un aumento de 11.8 %. Esto nos lleva que el error del modelo integrado y el caso base aumenta de forma homogénea tanto para el efecto del freno regenerativo y el consumo energético.



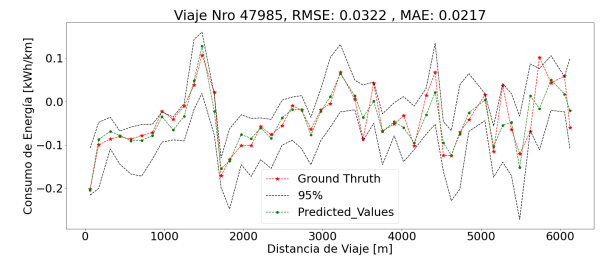
(a)



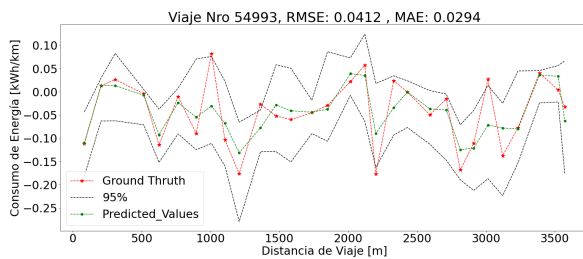
(b)



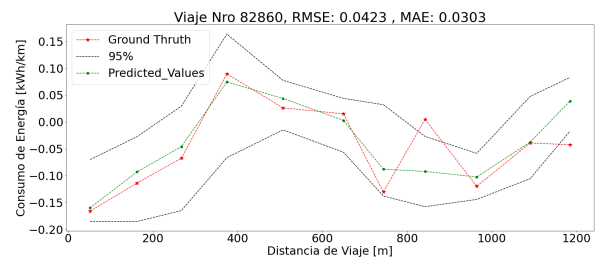
(c)



(d)



(e)



(f)

Figura 5.14: Media de las distribución para el intervalo de confianza y el valor esperado según la distribuciones de la figura ??

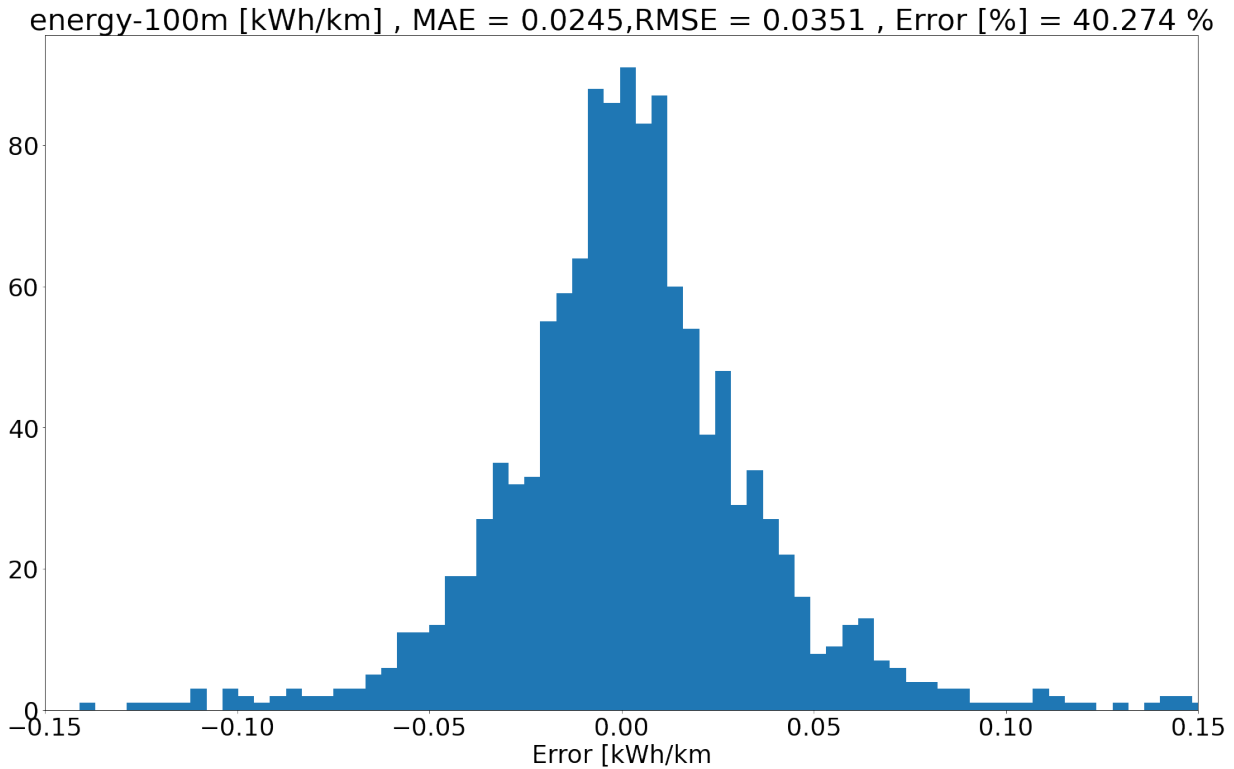


Figura 5.15: Error de modelo de consumo sobre todo el conjunto de validación, utilizando variable de velocidad estimada.

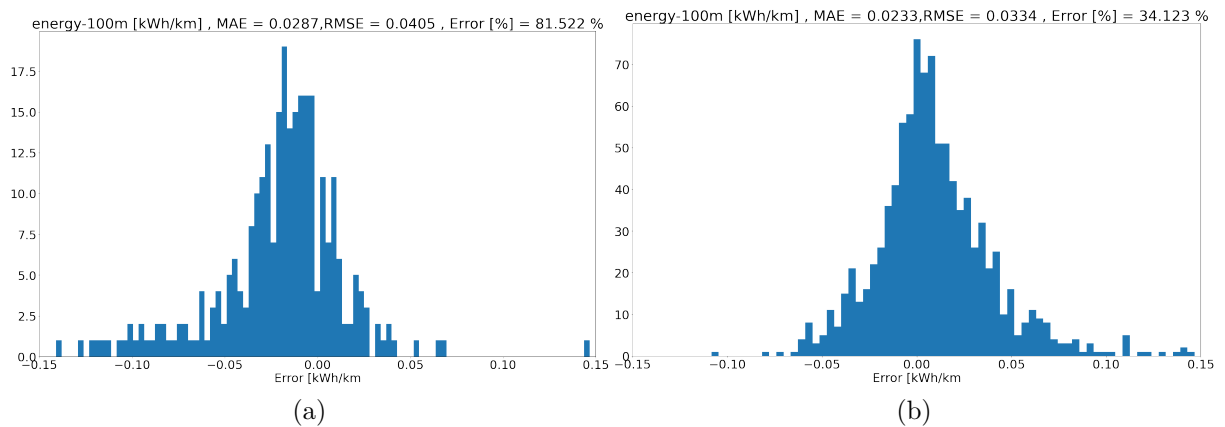


Figura 5.16: Error de estimación de consumo utilizando la velocidad estimada por el modelo anterior. (a) Distribución de error para freno regenerativo debido a los efectos del freno regenerativo (b) Consumo negativo debido a los cambios de velocidad positivos

Comparando la figura 5.17 y con la figura 5.11, se observa que en 4 de los 6 viajes existe una disminución el error asociado, pero de manera global se tiene que error aumenta de 7.8 % a 9.3 %, lo que se traduce en un aumento del MAE en 0.0027 kWh y de RMSE en 0.004 kWh, nos dice que el la media el error entre el modelo base y el modelo integrado es similar, esto

debido al buen desempeño del modelo de velocidad que permite estimar la velocidad con un error mínimo. Tomando también los valores de la media de la integral del consumo, es que tiene el valor real es de -6.91 kWh y el modelo estima en los 34 viajes un total de -6.45kWh, lo que se traduce en un error de 6.5%. Esto nos da una diferencia con el caso base, donde se sobre estimaba el consumo y en este caso, el modelo subestima el consumo total en 0.4 kWh. Esto nos muestra que en general el modelo de velocidad subestima el valor de salida, lo que lleva a que el modelo integrado subestime el consumo. Para esto se tiene como opción entrenar el modelo para estimar un quantile menor a 0.5, de esta forma se tiene un margen de seguridad, el que en general hará que el consumo sea mayor al estimado por la mediana.

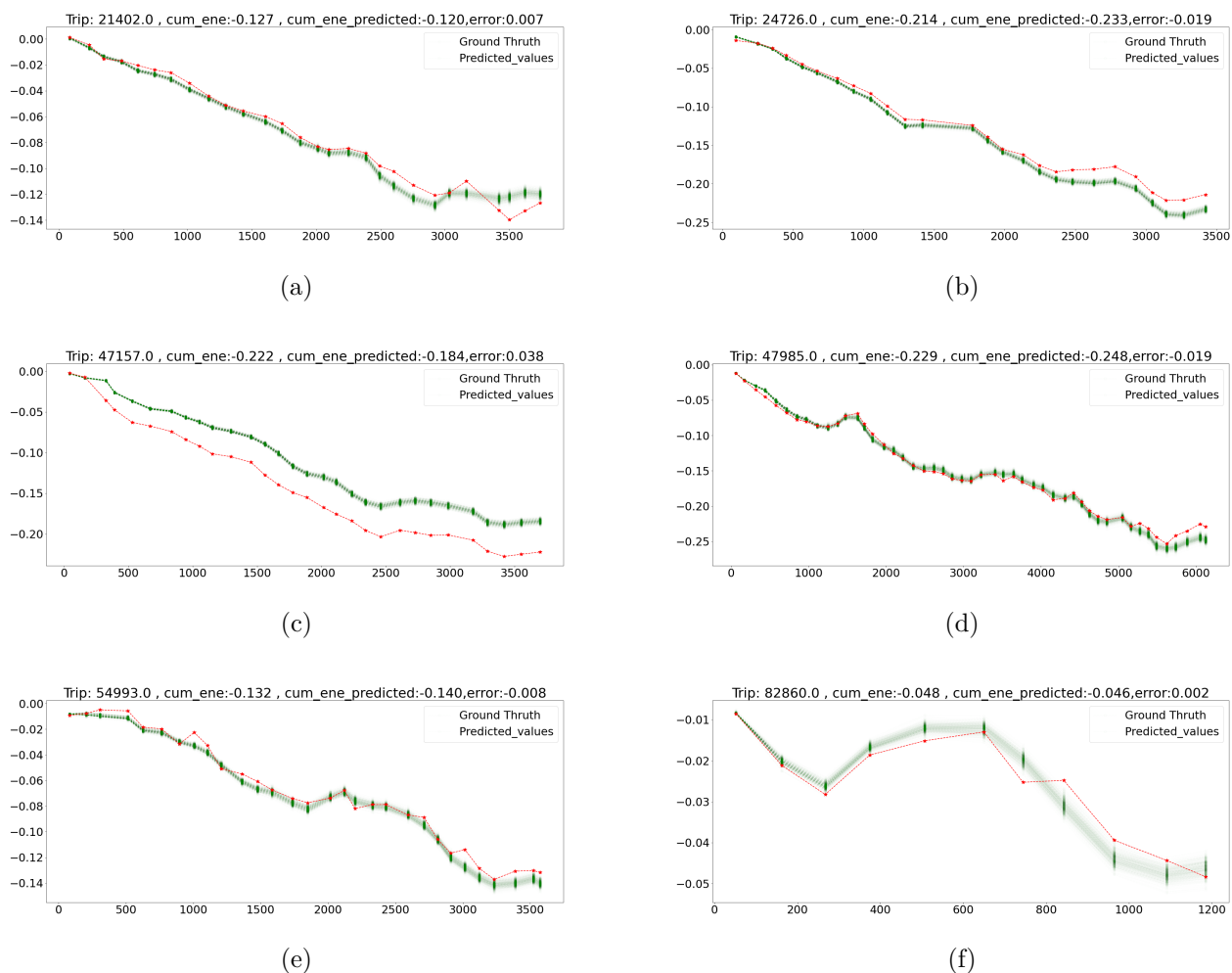


Figura 5.17: Integral del consumo esperado para la distribución obtenida en la figura ??, obteniendo la distribución del consumo total.

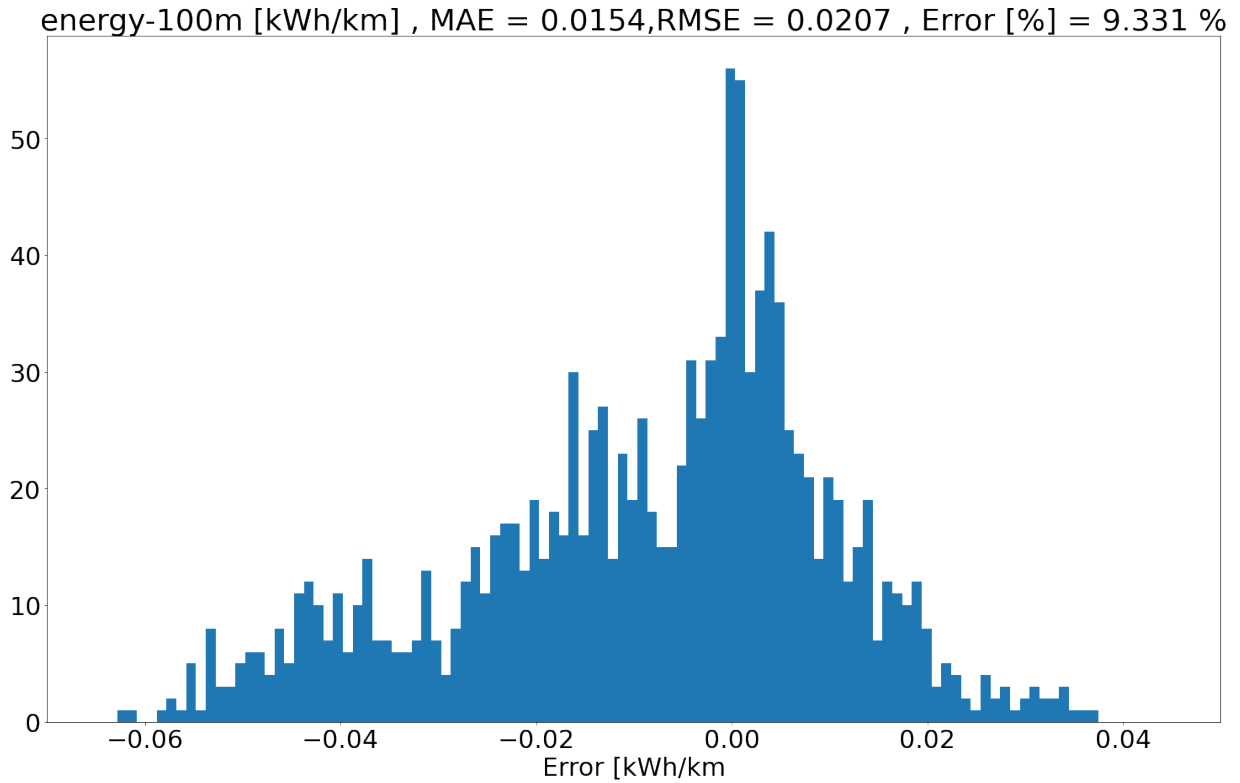
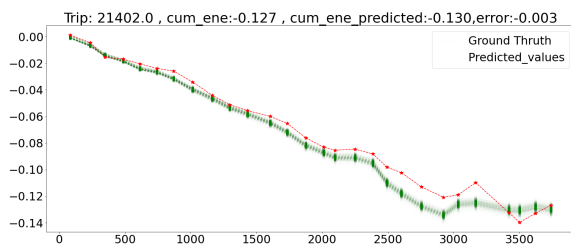


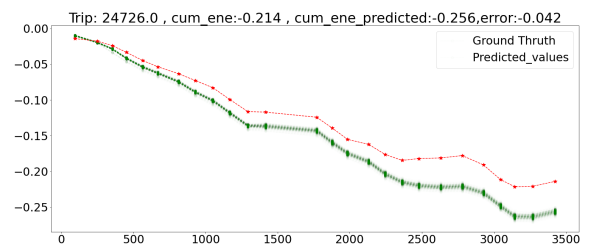
Figura 5.18: Error de modelo de consumo acumulado sobre todo el conjunto de validación, utilizando variable de velocidad rea y modelo entrenado con q0.5 (mediana)

5.3.1. Modelo de Sobre-Estimación de Consumo

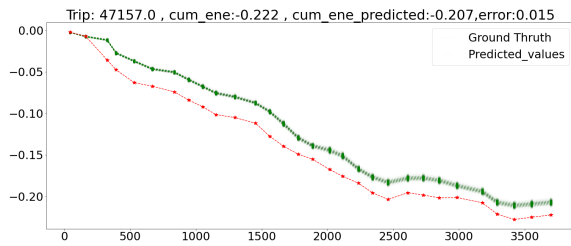
Dependiendo de la aplicación de este modelo, si se busca precisión estadística, el consumo debe ser estimado con el quantile 0.5, pero si se quiere aplicar por ejemplo para dar una recomendación si es posible llegar o no a destino, se debe tener en consideración que es más costoso que el vehículo quede detenido y no llegue al destino a que se recomiende cargar. Para solucionar esto es que se propone realizar la estimación con el quantile menor a 0.5, donde en este caso se utiliza el quantile 0.4 para realizar una comparativa entre los modelos. Los resultados obtenidos se muestran en la figura 5.19, donde se observa que en un aumento en gap de la integral de consumo, aumentando el error a una mayor sobre-estimación de este. Tomando el conjunto de validación completo se tiene un consumo total real de -6.91 kWh y para el consumo estimado es de -7.01 kWh o que se traduce en un 1% de sobre estimación, incluso si se compara con el error del modelo utilizando q0.5 de la figura 5.20, el error es menor, esto se debe a que el modelo de consumo no tiene mapeado todas las distribuciones que entrega el modelo de velocidad, esto se aprecia en los saltos de información que se ven en distintas distribuciones del modelo integrado, provocando que el modelo subestime el consumo global.



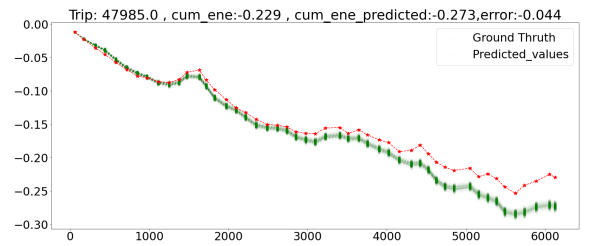
(a)



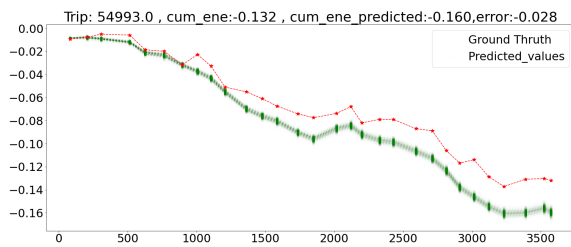
(b)



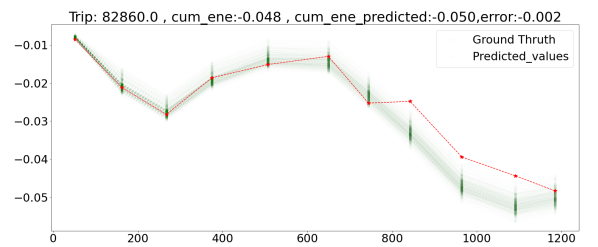
(c)



(d)



(e)



(f)

Figura 5.19: Resultados de en viajes seleccionados del conjunto de test, utilizando un modelo entrenado para estimar el cuantile 0.4

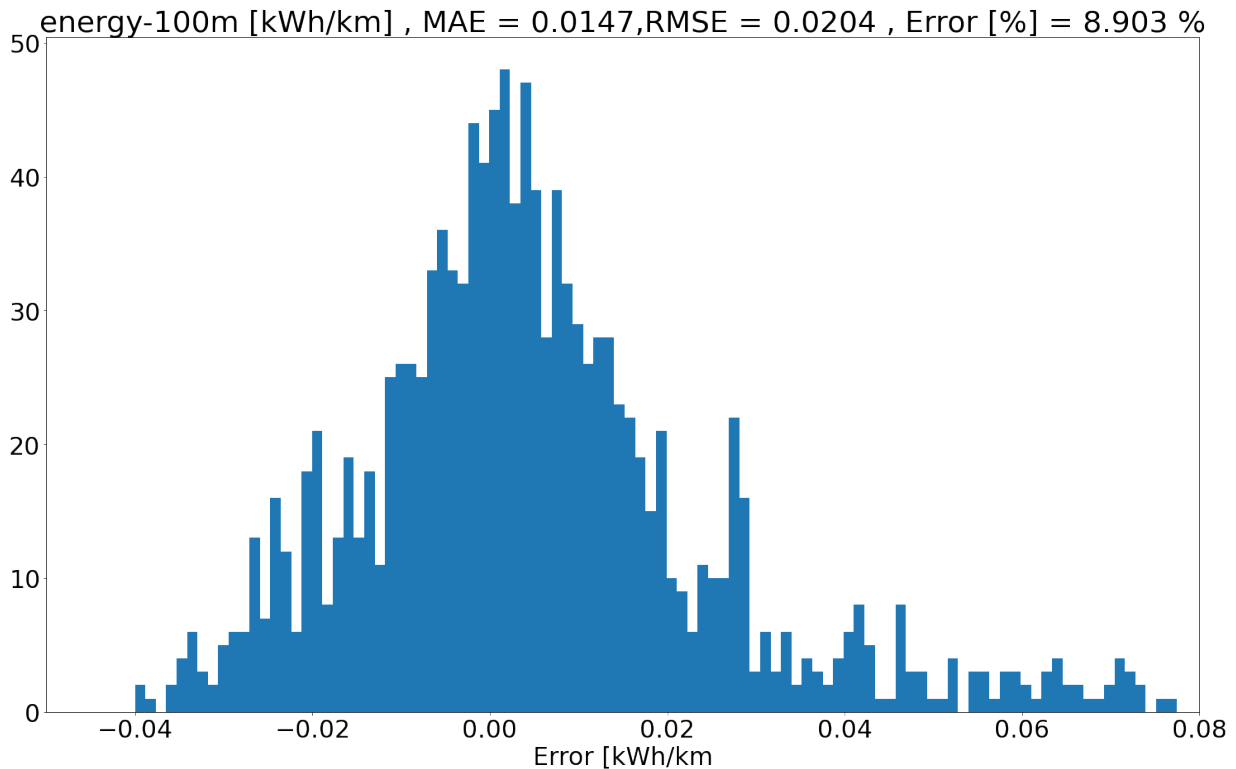


Figura 5.20: Error de modelo de consumo acumulado sobre todo el conjunto de validación, utilizando variable de velocidad real y modelo entrenado utilizando q0.4

Recapitulando las discusiones generales, se tiene que el modelo tiene un desempeño regular para estimaciones de el consumo por rango. llegando a un 40 % de error de acuerdo a la media de error global, esto nos dice que en promedio el modelo se equivoca un 40 % en cada estimación por rango, causando que no sea un buen recomendador señalando si ciertas secciones de la ruta requieres un gran consumo, lo que se puede traducir en una desconexión de batería por una baja de voltaje . Este problema se puede mejorar re-entrenando el modelo con data de otras zonas, donde existan distintos perfiles de consumo en pendiente, otra forma de mejorar el modelo es tomar variables que no se incluyeron en el análisis como el tipo de conductor, que permitirán generar un modelo (o variable de entrada) que caracteriza el estilo de conducción y por ende agregue una capa más la estimación de consumo.

Tomando el análisis de consumo por rango (consumo o regeneración), se tiene que el mayor error viene dado por los efectos de freno regenerativo, esto se puede mejorar tomando más variables del diagrama de Ishikawa de la figura 3.2, donde se tienen múltiples variables que afectan al freno regenerativo que no se están tomando en cuenta debido a que son variables medidas por la base de datos utilizada.

En relación al uso del modelo para estimar el consumo global de la ruta (integral sobre consumo por rango) se tiene que el modelo tiene una clara mejora con los estudios realizados en la sección 3.2.2.2, comparando el error del modelo q-0.5 presentado en la figura 5.18 y la tabla 3.4 se tiene un mejora de MAE en 83.2 % y una mejora de RMSE de 85.4 %. En relación

a la metodología propuesta en la figura 3.2.2.3, se tiene un mejora de 86.9%. Esto se debe a que el error asociado al modelo de consumo es gaussiano, por lo que se promedia al momento de integrarlo en la ruta, además se tiene que el buen desempeño del modelo de velocidad que permite que esta estimación tenga una mejora significativa a los paper existentes.

Capítulo 6

Conclusiones y Recomendaciones

A lo largo de este trabajo, se ha presentado la propuesta de metodología que se divide en 3 partes ,primero le modelo de estimación de distribución de velocidad, que es la variable con mayor impacto y mayor variabilidad, este modelo utiliza una red LSTM que tiene 5 variables de entrada, 2 variables que entregan información de los puntos anteriores y 2 que dan información del futuro. Este modelo tiene un error de 7.30 % lo que se traduce que en promedio el error es de 1 m/s o 3.6 km/hrs, esto es un error muy bueno, lo cual resultó clave el siguiente modelo. El siguiente modelo es el de estimación de consumo (caso base) que utiliza un árbol de regresión LightGBM, el cual tiene una mejora de velocidad de ejecución en comparación a XGBoost, manteniendo el desempeño de estimación, este modelo toma variables clave de la fenomenología de requerimientos de potencia, estimando la energía requerida en un rango de entre 50 y 100 metros. Este modelo tiene una error de un 27 %, lo que se traduce en un MAE de estimación de 0.0168 de error promedio en cada estimación, sin embargo se tiene que el error es mayor para la regeneración que el consumo, siendo casi 3 veces mayor. Para mejorar esto se requiere de mayor data que tenga información ya que esta corresponde al 23 % de la data total, por lo que no se logra representar completamente en el espacio de inferencia, la otra mejora es integrar el tipo de conductor, para definir la agresividad de conducción de este. A pesar de esto en la figura 5.12 se tiene una disminución del error de un 20 % los que se traduce en que el error del modelo es gaussiano, por lo que al integrar la estimación de consumo este error de promedia, aumentando considerablemente el desempeño en la estimación.

Unificando el modelo de estimación de distribución de velocidad y el modelo de estimación de consumo como se muestra en la figura 4.10, para el modelo de consumo q-0.5, esto resulta en un aumento del error de estimación en un 13 %, sin embargo, para la integral del consumo sobre las estimación, es decir, el consumo total estimado, se tiene un aumento del error de 1.5 % según el caso base. Esto deja en evidencia que se valida la hipótesis y que es posible estimar el consumo de un vehículo eléctrico de forma precisa, teniendo una mejora del 85 % sobre estudios realizados anteriormente.

Como recomendación para las futuras aplicaciones, es este modelo se puede utilizar para la optimización de flotas de vehículos y/o recomendación sobre si se llega al destino con cierta probabilidad. Este modelo será válido solo para zonas similares a Ann Arbor, Michigan,USA , es decir, zonas sin mucha inclinación de la ruta, por lo que si se quiere llevar este modelo a esos escenarios, el desempeño será drásticamente peor, pero se puede solucionar re-entrenando el

modelo de consumo y velocidad, ya que la metodología ha demostrado tener un muy buen desempeño en comparación a los artículos existentes. Para el modelo de velocidad se debe tomar un nuevo nodo correspondiente de zonas con alta inclinación para tener una estimación más certera. En el caso del modelo de consumo, se puede mejorar complejizando el modelo, tomando variables del el Diagrama de Ishikara de la figura 3.2, para lograr estimar los efectos del freno regenerativo con mayor exactitud, permitiendo así dar una recomendación tanto en relación al consumo total (integral de consumo por tramo), como dar alerta si algún tramo va a requerir un pick mayor consumo del que la batería puede entregar, como es el caso de una pendiente pronunciada.

6.1. Trabajos Futuros

De acuerdo a lo visto en el trabajo realizado, es que se propone como trabajo futuro, complementar la base datos utilizando información de otra zona, donde las características de las variables intrínsecas de la ruta sean diferentes, como por ejemplo data de Costa Rica, donde se tiene una gran diferencia de altitud entre las diferentes ciudades, de esta forma el modelo sea capaz de generalizar la salida, mejorando las estimaciones de velocidad y consumo. Para mejorar el desempeño general de la estimación de consumo, es que se debe estudiar a mayor profundidad las variables asociadas al freno regenerativo y consumo por efecto roce por rodadura, integrando el tipo de material de la calle transitada, de esta forma, se tiene caracterizado el coeficiente de roce entre neumático y el terreno, agregando una capa más que busca mejorar el desempeño en esta estimación y por ende el desempeño general de la herramienta, completando y complejizando el espectro del espacio de salida asociado al árbol de regresión, mejorando el desempeño de la herramienta bajo distintos escenarios.

Bibliografía

- Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., . . . et al. (2021, Dec). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76, 243–297. doi: 10.1016/j.inffus.2021.05.008
- Avcı, M. Y., Li, Z., Fan, Q., Huang, S., Bilgic, B., y Tian, Q. (2021). *Quantifying the uncertainty of neural networks using monte carlo dropout for deep learning based quantitative mri*.
- Cauwer, C. D., Verbeke, W., Coosemans, T., Faid, S., y Mierlo, J. V. (2017). A data-driven method for energy consumption prediction and energy-efficient routing of electric vehicles in real-world conditions. *Energies*, 10. doi: 10.3390/en10050608
- Chen, T., y Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Chen, Y., Wu, G., Sun, R., Dubey, A., Laszka, A., y Pugliese, P. (2020, 5). *A review and outlook of energy consumption estimation models for electric vehicles*.
- Chevalier, G. (2018). LARNN: linear attention recurrent neural network. *CoRR*, abs/1808.05578.
- Daniela, Trevor, H., Gareth, T. R. J., y Witten. (2021). *Tree-based methods*. Springer US. doi: 10.1007/978-1-0716-1418-1_8
- Edoardo, M., Ferenc, S., Stella, T., Daniel, G., y Georg, G. (2022). Shap-based explanation methods: A review for nlp interpretability anonymous acl submission.
- Fan, J., Ma, X., Wu, L., Zhang, F., Yu, X., y Zeng, W. (2019, 11). Light gradient boosting machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. *Agricultural Water Management*, 225. doi: 10.1016/j.agwat.2019.105758
- Felipe, J., Amarillo, J. C., Naranjo, J. E., Serradilla, F., y Díaz, A. (2015). Energy consumption estimation in electric vehicles considering driving style. En *2015 IEEE 18th International Conference on Intelligent Transportation Systems* (p. 101-106). doi: 10.1109/ITSC.2015.25
- Friedman, J. H. (2002). *Stochastic gradient boosting* (Vol. 38).
- Gal, Y., y Ghahramani, Z. (2016). *Dropout as a bayesian approximation: Representing model uncertainty in deep learning*.
- Gan, M., Pan, S., Chen, Y., Cheng, C., Pan, H., y Zhu, X. (2021, 5). Application of the machine learning lightgbm model to the prediction of the water levels of the lower columbia

- river. *Journal of Marine Science and Engineering*, 9. doi: 10.3390/jmse9050496
- Gómez-Gélvez, J. A., Mojica, C. H., Kaul, V., y Isla, L. (2016). *The incorporation of electric cars in latin america* (Inf. Téc.). Washington DC, Inter-American Development Bank.
- Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., y Schmidhuber, J. (2009). A novel connectionist system for unconstrained handwriting recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5), 855-868. doi: 10.1109/TPAMI.2008.137
- Irle, R. (2020). *Global plug-in vehicle sales reached over 3,2 million in 2020*.
- Javier Ochoa, . (s.f.). *Electromovilidad: ya circulan en chile 900 autos eléctricos*, <https://www.diarioconcepcion.cl/economia/2020/11/02/electromovilidad-ya-circulan-en-chile-900-autos-electricos.html>.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... Liu, T.-Y. (2017). *Lightgbm: A highly efficient gradient boosting decision tree*.
- Kim, E., Lee, J., y Shin, K. G. (2013). Real-time prediction of battery power requirements for electric vehicles. En *2013 acm/ieee international conference on cyber-physical systems (iccps)* (p. 11-20). doi: 10.1145/2502524.2502527
- Krzywinski, M., y Altman, N. (2017). Classification and regression trees. *Nature Methods*, 14, 757-758. doi: 10.1038/nmeth.4370
- Li, W., Stanula, P., Egede, P., Kara, S., y Herrmann, C. (2016). Determining the main factors influencing the energy consumption of electric vehicles in the usage phase. En (Vol. 48, p. 352-357). Elsevier B.V. doi: 10.1016/j.procir.2016.03.014
- Lundberg, S. M., y Lee, S. (2017). A unified approach to interpreting model predictions. *CoRR*, <abs/1705.07874>.
- Markets, y Markets. (2021). *Electric vehicle market report, region-global forecast 2030*, <https://www.marketsandmarkets.com/market-reports/electric-vehicle-market-209371461.html>.
- Oh, G. S., Leblanc, D. J., y Peng, H. (2019). *Vehicle energy dataset (ved), a large-scale dataset for vehicle energy consumption research*.
- ONU Environment Programme, M. (2019). *Electric mobility: Status in latin america and the caribbean and opportunities for regional collaboration 2019* (Inf. Téc.). ONU.
- Pevec, D., Babic, J., Carvalho, A., Ghiassi-Farrokhfal, Y., Ketter, W., y Podobnik, V. (2019). Electric vehicle range anxiety: An obstacle for the personal transportation (r) evolution? En *2019 4th international conference on smart and sustainable technologies (splitech)* (pp. 1-8).
- Rietmann, N., Hügler, B., y Lieven, T. (2020). Forecasting the trajectory of electric vehicle sales and the consequences for worldwide co2 emissions. *Journal of Cleaner Production*, 261, 121038.
- Rozas, H., Troncoso-Kurtovic, D., Ley, C. P., y Orchard, M. E. (2021, 8). Lithium-ion battery state-of-latent-energy (sole): A fresh new look to the problem of energy autonomy prognostics in storage systems. *Journal of Energy Storage*, 40. doi: 10.1016/j.est.2021.102735
- Shapley, L. S. (1952). *A value for n-person games*. Santa Monica, CA: RAND Corporation.

doi: 10.7249/P0295

Weatherspark. (2022). *El clima en ann arbor, el tiempo por mes, temperatura promedio (michigan, estados unidos) - weather spark*, <https://es.weatherspark.com/y/16511/clima-promedio-en-ann-arbor-michigan-estados-unidos-durante-todo-el-a%C3%B1o>.

Zhang, J., Wang, Z., Liu, P., y Zhang, Z. (2020, 10). Energy consumption analysis and prediction of electric vehicles based on real-world driving data. *Applied Energy*, 275, 115408. doi: 10.1016/j.apenergy.2020.115408

Anexo A

Ficha Técnica EV Nissan Leaf 2013

Tabla A.1: Especificaciones técnicas Nissan Leaf 2013

Prestaciones y consumos homologados	
Velocidad máxima	145 km/h
Aceleración 0-100 km/h	11,9 s
Dimensiones, peso, capacidades	
Tipo de Carrocería	Turismo
Número de puertas	5
Longitud	4.450 mm
Anchura	1.770 mm
Altura	1.550 mm
Batalla	2.700 mm
Vía delantera	1.535 mm
Vía trasera	1.535 mm
Coefficiente Cx	0,29
Superficie frontal	2,27 m ²
Factor de resistencia	0,66
Peso	1.600 kg
Volúmenes de maletero	
Volumen mínimo con dos filas de asientos disponibles	330 litros
Número de plazas	5
Distribución de asientos	2 + 3
Motor Eléctrico	
Propósito	Impulsar al vehículo / generar corriente eléc
Potencia máxima	109 CV / 80 kW
Revoluciones potencia máxima	9.800 rpm
Par máximo	280 Nm
Revoluciones par máximo	2.730 rpm
Ubicación	Delantero transversal
Tensión nominal	360 V
Batería	
Tipo	Acumulador de iones de litio
Ubicación	Central trasero
Capacidad	24 kWh

Anexo B

Estructura DataFrame

Tabla B.1: Estructura de dataframe, el cual es normalizado para el posterior entrenamiento del modelo de estimación de velocidad.

Hora	Distancia de Viaje [m]	Distancia a Nodo [m]	Distancia a Siguiente Calle [m]	Precipitación [mm]	Humedad [%]	Visibilidad [%]	Nieve [mm]
14	0	1981	1649	0	64	10	0
14	56	1925	1593	0	64	10	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
11	1540	0	0	5	44	12%	0
Disco Pare	Ceda el Paso	Paso Peatonal	Calle Primaria	Calle Secundaria	Calle Terciaria	Calle Residencial	Inclinación [°]
1	0	0	0	1	0	0	0.01
1	0	0	0	1	0	0	0.02
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0	0	0	0	0	1	0	0.1

Anexo C

Modelo de Red LSTM

Tabla C.1: Modelo de red LSTM utilizado para el proceso de estimación de Velocida

Layer (type)	Output Shape	Param #
cu_dnnlstm_2 (CuDNNLSTM)	(None, 32)	6144
repeat_vector_1 (RepeatVector)	(None, 5, 32)	0
cu_dnnlstm_3 (CuDNNLSTM)	(None, 5, 64)	25088
monte_carlo_lstm_1 (MonteCarloLSTM)	(None, 5, 128)	98816
time_distributed_1 (TimeDistributed)	(None, 5, 1)	129
Total params: 130,177		
Trainable params: 130,177		
Non-trainable params: 0		

Anexo D

Análisis Proceso de MonteCarlo

Tabla D.1: Análisis de nro de iteraciones en proceso de Monte Carlo

Nº procesos Monte Carlo	Mean [m/s]	STD [m/s]	q25[m/s]	q50[m/s]	q75[m/s]
10	10.817	5.514	6.772	11.425	15.259
20	10.817	5.515	6.769	11.416	15.264
30	10.816	5.515	6.767	11.421	15.256
40	10.817	5.514	6.760	11.420	15.262
50	10.816	5.514	6.762	11.423	15.263
60	10.817	5.515	6.764	11.422	15.262
70	10.816	5.515	6.760	11.419	15.261
80	10.817	5.514	6.765	11.422	15.261
90	10.817	5.514	6.761	11.418	15.263
100	10.817	5.514	6.767	11.419	15.260
125	10.817	5.514	6.759	11.421	15.261
150	10.817	5.514	6.767	11.419	15.261
200	10.817	5.514	6.768	11.422	15.261
300	10.817	5.515	6.761	11.420	15.261
500	10.817	5.514	6.763	11.420	15.261
750	10.817	5.514	6.765	11.420	15.260
1000	10.817	5.514	6.765	11.420	15.260

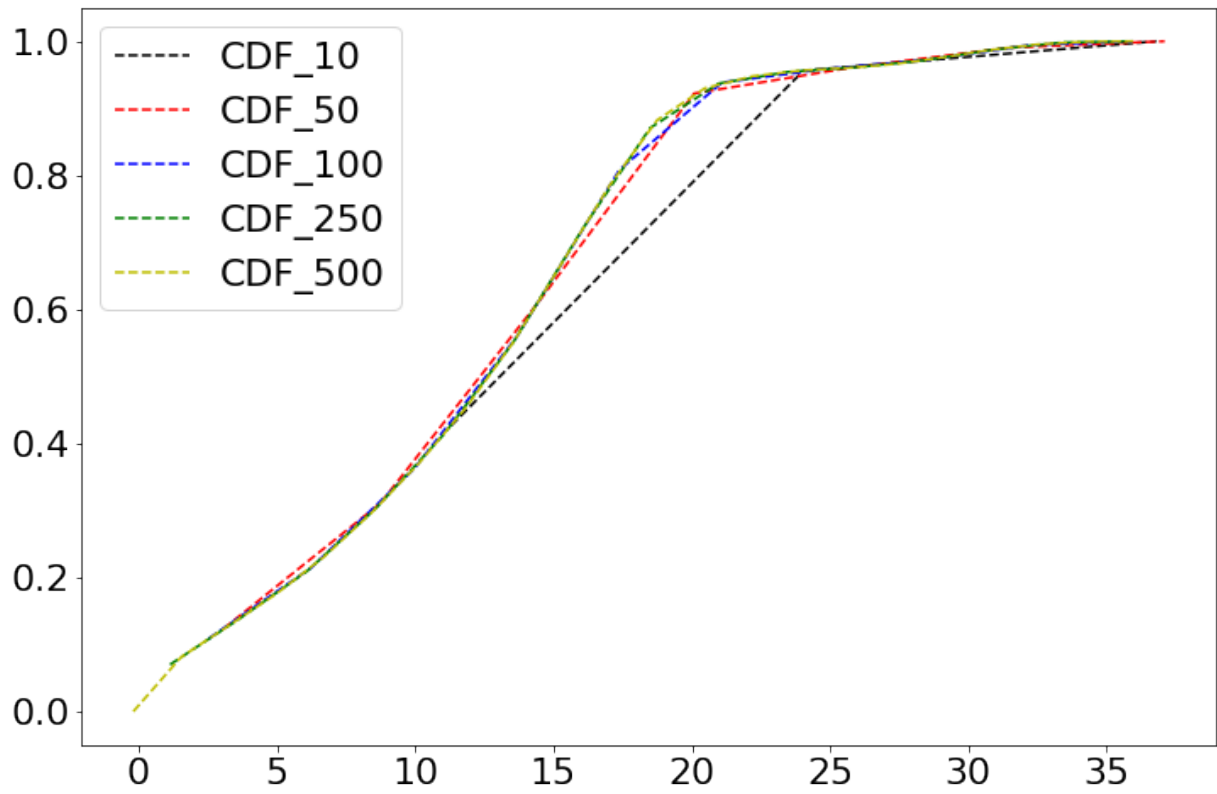


Figura D.1: CDF de la distribución de velocidades utilizando distintos número de procesos de Monte Carlo

Anexo E

Correlación Entre Periféricos y Condiciones Ambientales

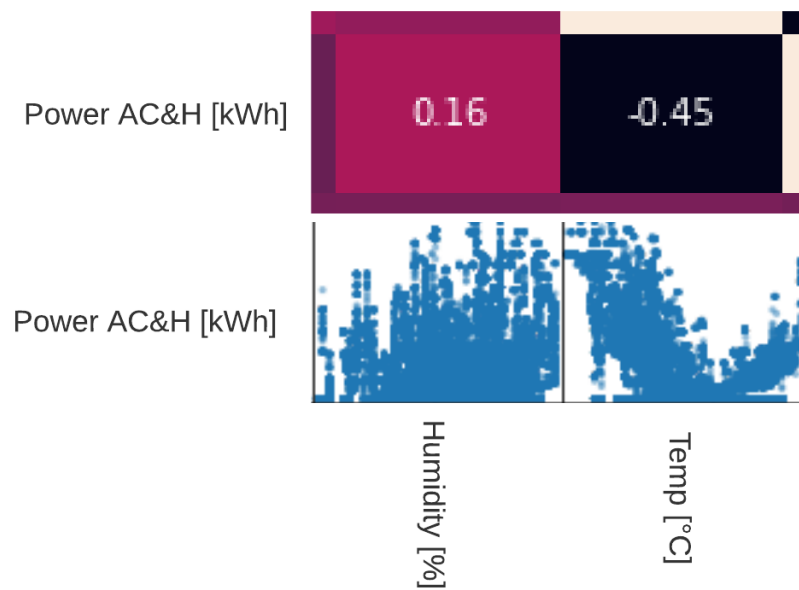


Figura E.1: Correlación entre consumo de aire acondicionado y calefacción y variable humedad y temperatura.