# MARCO TEÓRICO

1.1 Machine learning

1.1.1 Fake news

El concepto o definición de noticias falsas puede tener diferentes significados en diferentes contextos, la cual puede generar confusión en algunos lectores, es por ello que es necesario para la presente investigación aclarar el concepto que se tomara de fakenew una “fakenew” además de otros conceptos adicionales que serán usados en la investigación. Por un lado, la palabra mencionada “falso” o en inglés “fake” se entiende como algo que no es cierto o aparenta ser verdadero, pero no lo es. Por otro lado, la palabra mencionada “noticias” o en inglés “news”, se refiere a un informe de sucesos del presente o del futuro. Por lo tanto, el enunciado “noticias falsas”, indica una información falsa divulgada en forma de hechos verídicos por medio digital o público. Asimismo, cabe destacar que, en Malasia, está prohibido la divulgación de noticias falsas, la cual si se infringe la legislación obtendrá una sanción de hasta seis años de cárcel(Mokhtar et al. 2019). En base a esta definición de fake news se puede deducir que son noticia perjudiciales que fomentan la desinformación y el desorden así mismo traer consecuencias nefastas es por ello que a lo largo de los años se han ido desarrollado softwares capaces de filtrar esas noticias falsa mediante el aprendizaje automático, la cual es una disciplina científica de la rama de inteligencia artificial, que se encargan de crear sistemas que aprenden automáticamente tales como relación de noticias, transcripción de voz a texto, publicidad a los usuarios en cierta red, selección de resultados de búsquedas relevantes e identificación de objetos a través de imágenes (Lecun, Bengio, and Hinton 2015).

1.1.2 Aprendizaje automático

El aprendizaje automático tiene diversos modelos o algoritmos matemáticos que nos permiten la representación de datos que en su mayoría se clasifican en aprendizaje supervisado y no supervisado, los cuales se diferencia en que para los modelos de aprendizaje supervisado se conoce la variable f que permite asignar labels (etiquetas) a los datos, mientras que en el otro caso el aprendizaje no supervisado se basa en un entrenamiento conjunto de vectores y no se cuenta con labels (etiquetas) asignados a los datos para el entrenamiento dificultando la agrupación de datos.(J. Nilsson 1996)

1.2.2.1 Regresión logística

La regresión logística es un tipo de aprendizaje supervisado que pertenece al tipo de análisis basado en la regresión que es usado para poder predecir el resultado de una variable respecto a las otras variables que son predictoras o independientes. Asimismo, la regresión logística incorpora o incluye la comprobación de hipótesis, en conjunto con las demás evaluaciones realizadas, medidas de aptitud o cálculos. Además, cabe resaltar que la regresión logística podría ser unos de los métodos de explotación en lo que respecta a los datos, los más habituales o usados, ya que cuenta con semejanza o analogía matemática con la sencillez en la que se selecciona de manera automática las variables y la regresión múltiple ordinaria (Desouky Fattoh and Mousa 2021).

1.2.2.2 Bosque aleatorio

El clasificador de bosque aleatorio es otro tipo de aprendizaje automático supervisado que a diferencia de la regresión logística está compuesto por una gran variedad de árboles de decisiones, las cuales se desempeñan de manera óptima como un conjunto. Asimismo, por cada árbol individual que se fabrica o construye, realiza uso de la aleatoriedad, para que pueda establecer un bosque de árboles relacionados de forma negativa, la cual su predicción para cualquier árbol será la más confiable. Además, se asigna a cada árbol en el bosque aleatorio una prevención de clase, para lo cual el modelo que se convertirá en predictor será la clase más votada. Cabe destacar, que funciona bien el modelo de bosque aleatorio, ya que los modelos constituyentes individuales se superan gracias a la gran cantidad de árboles correlacionados. Por lo tanto, los bosques aleatorios es un procedimiento o método de vital importancia porque se encuentra basado en el algoritmo de árboles de clasificación y regresión (Desouky Fattoh and Mousa 2021).

1.2 Aprendizaje profundo

El aprendizaje automático es un tipo de machine learning que se diferencia principalmente en que el Deep learning se basa en un proceso de aprendizaje automático usando redes neuronales, las cuales resultan muy útiles para la revelación de una o varias estructuras relacionadas con datos de dimensión alta, que a su vez son algoritmos aplicados a diversos dominios tales como los negocios en internet, la ciencia, entre otros (Lecun et al. 2015), mientras que en el machine learning se usa usualmente algoritmos que permite la toma de decisiones mediante ramas de evolución o arboles de decisión, esta diferencia entre ambos implica que el aprendizaje automático este sujeto a la supervisión humana mientras que el aprendizaje profundo sea más autónomo. Así mismo en el aprendizaje profundo mencionaremos 2 tipos de redes neuronales más comúnmente usadas en el aprendizaje profundo las cuales son CNN (Convultional Neural Networks), RNN (Recurrent Neural Networks).

1.2.1 Redes neuronales convolucionales

Es un algoritmo de aprendizaje profundo, la cual tiene la capacidad como entrada una imagen ya sea en 2D o 3D, para posteriormente procesarla y al final poder diferenciarlas entre otras imágenes(Lecun et al. 2015).

1.2.2 Redes neuronales concurrentes

Las redes neuronales concurrentes son usadas para procesar en su entrada una secuencia de elementos a la vez, conteniendo información de elementos pasados basados en historias pasadas, a menudo se usa las redes neuronales recurrentes para tareas relacionadas con la entrada de datos secuenciales, tales como el lenguaje verbal y texto(Lecun et al. 2015), permitiendo identificar diversas características del lenguaje con en el caso de Google translate que utiliza el procesamiento de lenguaje natural o NLP, además las redes neuronales concurrentes poseen una característica de memoria a corto plazo (LSTM).

1.2.2.1 Memoria a corto plazo (LSTM)

La arquitectura de red neuronal recurrente artificial LSTM o memoria a corto plazo, tiene la capacidad de que a través de una larga duración pueda registrar patrones de forma selectiva. Para modelar datos secuenciales puede ser una excelente alternativa, así como también para el estudio de dinámica complejas en lo que respeta al comportamiento humano. Por lo tanto, LSTM o memoria a corto plazo se utiliza mayormente en aplicaciones de múltiples modelos, tales como subtitulado de fotografías tomadas, asimismo también realiza mejoras en las redes neuronales recurrentes o RNN, en contextos híbridos(Desouky Fattoh and Mousa 2021).

1.1.4. Procesamiento del lenguaje natural (NLP)

El procesamiento de lenguaje natural pertenece a una parte de la inteligencia artificial enfocada a la arquitectura de red neuronal recurrente artificial LSTM , la cual se basa en el análisis y procesamiento de un conjunto de datos respecto al lenguaje natural, como puede ser el lenguaje verbal y escrito. Para el caso de la escritura o texto, el propósito es que el texto no estructurado sea cambiado hacia una manera un poco más estructurada, para que pues e pueda obtener información útil cuando se realiza el análisis. Asimismo, el procesamiento del lenguaje natural abarca una serie de metodologías o técnicas para organizar y realizar el procesamiento de los datos de sentimientos, las cuales pueden ser módulos lingüísticos, vectorización de texto o tokenización(Hybrid, Hawks, and Feature 2021).

1.1.5 Módulos lingüísticos y de tokenización

Es el proceso por el cual se fraccionan en partes más reducidas el texto, tales como en palabras o expresiones(Hybrid et al. 2021).

1.1.6 Vectorización de texto

Es un proceso, la cual el texto no estructurado se convierte o transforma en un resultado más organizado, tales que se puedan aplicar los algoritmos de aprendizaje automático con metas a una extracción de conocimientos o de minería, es decir se extraen u obtienen las particularidades numéricas empleando cálculos estadísticos que están basadas en texto o palabras(Hybrid et al. 2021).

1.1.7 Máquina de vectores de soporte (SVM)

La máquina de vectores de soporte SVM o también conocido por sus siglas SVM, son una agrupación de algoritmos de aprendizaje supervisado, es decir SVM se puede utilizar como un clasificador de pares o también como para la clasificación múltiple. Asimismo, actúa aplicando una técnica para muchas clases en contra de una por cada par de clases. Por lo tanto, el propósito de la máquina de vectores de soporte es identificar o reconocer las instancias, la cual tiene dependencia a una ecuación lineal, así también se puede ejecutar con una función de kernel, una clasificación no lineal(Desouky Fattoh and Mousa 2021).