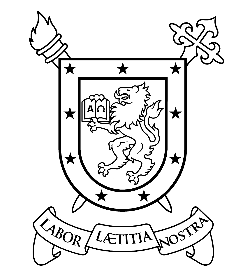
**UNIVERSIDAD DE SANTIAGO DE CHILE**

**FACULTAD DE INGENIERIA**

**DEPARTAMENTO DE INGENIERIA INFORMATICA**

**PREDICCIÓN DE ENERGÍA EÓLICA CON MACHINE LEARNING**

Propuesta de Trabajo de Título para Ingeniero Civil en Informática

|  |  |
| --- | --- |
| Nombre:  R.U.N.: | Manuel Moya Muñoz 14.332.262-9 |
| Año Ingreso: | 2017 |
| Teléfono:  E-mail:  Profesor Guía: | +56 9 8277 5722 manuel.moya.mu@usach.cl Gonzalo Acuña |

01 de julio de 2019

**Contenido**

[**1** **Objetivos del proyecto** 3](#_Toc14698165)

[**1.1** **Objetivo general** 3](#_Toc14698166)

[**1.2** **Objetivos específicos** 3](#_Toc14698167)

[**2** **Descripción del problema** 3](#_Toc14698168)

[**2.1** **Motivación** 3](#_Toc14698169)

[**2.2** **Estado del arte** 4](#_Toc14698170)

[**2.3** **Definición del problema** 7](#_Toc14698171)

[**3** **Definición de la solución** 7](#_Toc14698172)

[**3.1** **Alternativas de solución** 7](#_Toc14698173)

[**3.2** **Características de la solución** 8](#_Toc14698174)

[**3.3** **Propósito de la solución** 8](#_Toc14698175)

[**3.4** **Alcances y limitaciones de la solución** 9](#_Toc14698176)

[**4** **Metodología y herramientas** 9](#_Toc14698177)

[**4.1** **Metodología** 9](#_Toc14698178)

[**4.2** **Herramientas** 10](#_Toc14698179)

[**4.3** **Plan de trabajo** 11](#_Toc14698180)

[**5** **Referencias** 11](#_Toc14698181)

1. **Objetivos del proyecto**

## **Objetivo general**

Generar un modelo predictor de energía eólica de corto plazo basado en Machine Learning (ML) logrando mejores predicciones con los datos disponibles. El conocimiento extraído debe ser aplicable en los parques eólicos ubicados en Chile, de esta forma aportar a la industria nacional generando sistemas de predicción de energía eólica de más bajo costo.

## **Objetivos específicos**

* Generar un marco de trabajo que permita entender, describir y modelar los datos, para abordar el problema de predicción de energía eólica.
* Crear un modelo predictor, de corto plazo, de producción de energía eólica utilizando técnicas de ML.
* Establecer un cuadro comparativo del resultado del modelo predictor en base al análisis de información estadística disponible (línea base) y otras técnicas de ML descritas en la literatura académica con datos similares.
* Desarrollar un sistema de predicción en base al modelo creado y al conjunto de datos utilizado, que pueda ser empleado como herramienta complementaria en un parque eólico.

1. **Descripción del problema**

## **Motivación**

Este proyecto de título conjuga la problemática de la predicción de la energía eólica y las potenciales soluciones que entrega la Inteligencia Artificial (IA), específicamente con el uso de técnicas de Machine Learning (ML), catalogada en el 2018 como componente tecnológico relevante [1]. La motivación se describe en los siguientes puntos:

* La energía que se genera actualmente en el sistema eléctrico de Chile (gran escala), es difícil de almacenar [2], por lo tanto, la planificación de producción de un parque eólico requiere información certera, la más cercana a la realidad.
* El sistema eléctrico debe poder asegurar la demanda, así mantener el balance con la oferta, de esta forma evitar el colapso del sistema. Por lo anterior es fundamental predecir valores de los parques eólicos e integrar esta energía de forma eficiente al sistema eléctrico nacional [2].
* Se necesita predicción de energía eólica para planificar producción, tomar decisiones de incrementarla y/o disminuirla, incluyendo decisiones económicas; adicionalmente tomar medidas de seguridad energética cuando sean requeridas [2].
* Modelos estadísticos tradicionales (basados en datos de modelos atmosféricos) tienen porcentaje de error mayor a modelos basados en series de tiempo [3] y algunos no reflejan de buena forma la realidad.
* Aumento de Energía Renovable No Convencional (ERNC) a través de Plan Energía 2050 de Chile [4], el cual tiene como meta el 70% de ERNC para el 2050.
* La predicción de energía eólica es compleja, ya que depende de condiciones meteorológicas, la cuales son muy variables.

## **Estado del arte**

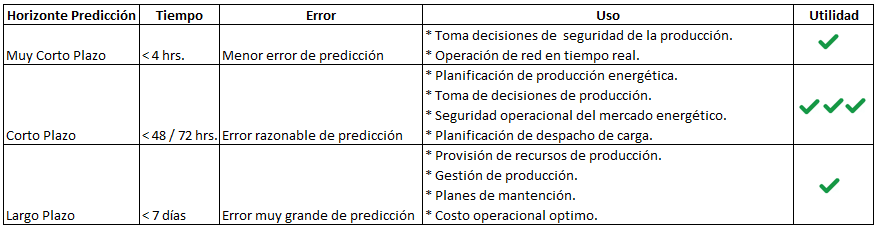
El uso industrial del ***viento en Chile*** inició con la molienda de cereales y en las salitreras para bombear agua en las primeras décadas del siglo XX. En cuanto a la generación de energía eólica, su uso comenzó en el año 2001. En las zonas costeras los vientos son ideales para la generación de energía, por lo que existe un gran potencial en el país para esta industria [5].

Actualmente, la ERNC en Chile es del 24%, donde la ***energía eólica*** ocupa el 6,02%, lo que equivale a la generación de 1540 MW. Adicionalmente para este tipo de energía se estiman proyectos por 500 MW/año (aproximadamente). Esto clasifica a Chile como el segundo país en capacidad instalada de energía eólica en América Latina y el Caribe [6]. En relación con la industria de ERNC, Chile ha establecido la Política Energía Nacional 2050, la cual pone como meta tener un 70% de ERNC en el 2050, lo que incrementará la generación de energías renovables incluyendo la eólica.

El proceso de ***generación de energía eólica*** consta en convertir la energía cinética del viento en energía eléctrica mediante turbinas que están diseñadas para aprovechar el efecto Venturi y realizar la generación de energía. En cuanto a los parques eólicos actuales, usan turbinas modernas de 2 a 3 MW, permitiendo producir la energía de forma eficiente.

En ML la forma de ***predicción*** más usada es vía series de tiempo, es decir, con datos de entrada (muestras) tomados a intervalos de tiempo periódicas o regulares. Las series de tiempo tienen dos restricciones, dependen del tiempo y tienen algún tipo de estacionalidad. Es importante señalar que la predicción de energía eólica se divide en tres grupos, según su horizonte de predicción y cada uno de ellos tiene un uso distinto considerando el respectivo error asociado, como muestra la Tabla 1.

Tabla 1: Horizontes de predicción energía eólica [3].

******

La predicción de energía eléctrica y más aún cuando es eólica, tiene una complejidad intrínseca debido a su incertidumbre de predicción, esto debido a que la naturaleza de la meteorología es explicada por la dinámica de fluido y el comportamiento caótico de la atmosfera [7], esto conlleva un grado de error en las estimaciones estadísticas y físicas (usando modelos de leyes de la física para predecir energía eléctrica). Esta incertidumbre también se puede visualizar en Figura 1, es decir, la potencia generada a 10 m/s es 8 veces mayor la generada a 5 m/s. Un error en la predicción puede significar un alto error en la generación de energía eléctrica.

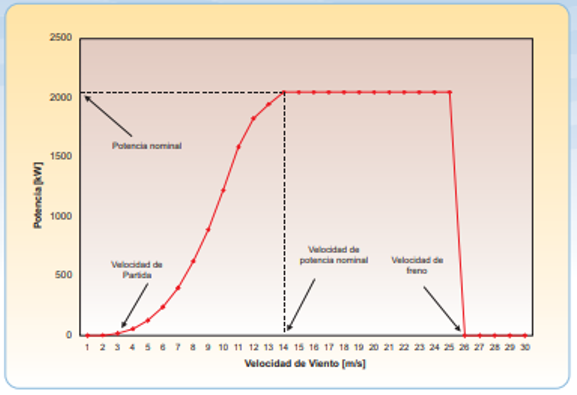


Figura 1: Curva de potencia de aerogenerador de 2 MW [8].

***Machine Learning (ML)*** o aprendizaje de máquina, como definió Arthur Samuel, pionero en inteligencia artificial de IBM y Stanford: “es el campo de estudio que da a las computadoras la habilidad de aprender sin ser explícitamente programadas” [9], su objetivo es detectar patrones de forma automática a partir de datos de entradas. Los algoritmos empleados se ejecutan en un proceso de aprendizaje supervisado o no supervisado. Como se puede ver en la Figura 2, ML es una subárea de inteligencia artificial, actualmente de forma errónea son usadas como sinónimos. A su vez ML tiene como subárea todo lo relacionado Deep Learning. ML es un nuevo campo de la ingeniería informática, la cual ha tenido un auge en los últimos años. Sus algoritmos están basados principalmente en redes neuronales artificiales y las soluciones son abordadas a través de clasificación o regresión [10]. Su utilización ha tomado fuerza, siendo en el 2018 declarado componente tecnológico relevante por la Association for Computing Machinery (ACM).

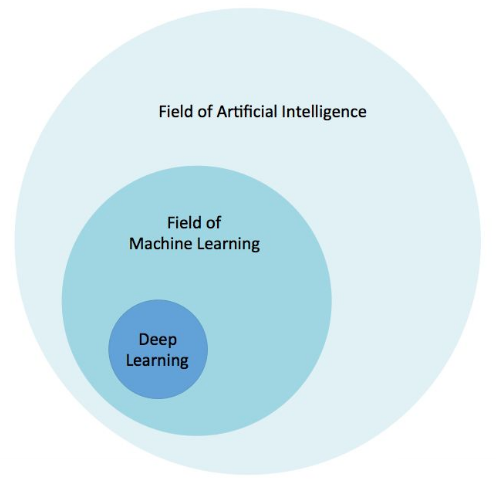


Figura 2: Alcance de ML [9].

## **Definición del problema**

Aunque la predicción del viento ha mejorado con el tiempo, aún tiene un alto porcentaje de error debido a su naturaleza caótica, lo cual genera intermitencias, estas intermitencias hacen que la generación de energía eólica tenga muchas fluctuaciones [11]. Esto provoca problemas en la integración de la energía eólica al sistema eléctrico chileno, ya que dificulta el trabajo de operadores eléctricos y su planificación, haciéndose complejo determinar la producción necesaria para cubrir la demanda y pudiendo generar problemas de seguridad del sistema eléctrico. Este problema se irá acrecentando a medida que se vaya cumpliendo el Plan Energía 2050, el cual contempla un crecimiento del 70% de las ERNC en Chile.

1. **Definición de la solución**

## **Alternativas de solución**

Los *modelos físicos* están basados en variables físicas atmosféricas y condiciones físicas de las plantas eólicas, usan servicios meteorológicos actuales para predecir mediante sistemas matemáticos complejos. Los *modelos estadísticos* sebasan en modelos numéricos para realizar las predicciones, el más utilizado es la predicción mediante regresión lineal. También existe un grupo de técnicas *Machine Learning* para generar la predicción, estas son: Support Vector Machine, Adaptative Bayesian Learning, K-means Clustering con redes neuronales [12] y Extreme Learning Machine (ELM) [13]. En este grupo están las redes neuronales artificiales, las cuales toman relevancia, ya que reportan mejoras de al menos de un 13% con respecto a las mediciones existentes [14].

## **Características de la solución**

Esta solución será basada en redes neurales artificiales ya que está comprobada su eficiencia con respecto a mediciones existentes, además es un componente tecnológico relevante que toma cada vez mayor preponderancia. Para la solución inicialmente se hará una investigación para entender los conceptos de predicción de energía eólica a corto plazo, también lo relacionado a redes neuronales artificiales, abordando puntos como recolección y entendimiento de datos, diseño neuronal (perceptrón multicapa y Long Short Term Memory -LSTM-) y entrenamiento de la red. Posteriormente se implementará el modelo con sus algoritmos de redes neuronales artificiales, se harán mediciones de error del modelo, se comparará con datos existentes en parque eólico y otras técnicas de ML disponibles desarrolladas con datos similares. Los datos principales del modelo son fecha, hora, velocidad del viento y energía generada como se visualizan en Figura 3. Adicionalmente se pueden considerar datos como dirección del viento y densidad del aire.

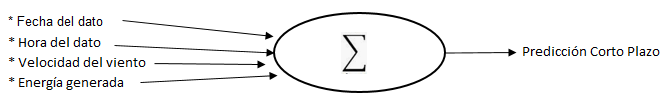


Figura 3: Datos del modelo [15].

Estos datos utilizados, serán de un parque eólico chileno, si esto no es posible se usará un conjunto de datos disponible en el sitio Kaggle.com (sitio de competencia de ML con datos gratis disponibles para su uso), de esta forma crear el modelo predictor. En ambos casos el conocimiento generado será aplicado a la realidad chilena.

## **Propósito de la solución**

Generar un modelo predictor novedoso y de bajo costos basado en ML, que permita a los parques eólicos nacionales obtener predicciones más certeras con una menor tasa de error, así realizar de buena forma la planificación de la producción energética para cumplir la demanda, apoyar la toma de decisiones de producción, abastecimiento y económica de un parque eólico. Además, poder evitar problemas de seguridad de la matriz energética, reduciendo el impacto de potenciales pérdidas económicas en el sistema eléctrico chileno.

## **Alcances y limitaciones de la solución**

Los alcances y limitaciones de este proyecto de título son:

* El resultado del modelo ML, será comparado con predicciones estadísticas que posea el parque eólico y con resultados de otros modelos de ML disponibles en literatura académica con datos similares.
* Solo se creará un modelo predictor utilizando técnica de ML.
* Se utilizará un conjunto de datos de un parque eólico chileno, en caso de no obtener estos datos, se utilizarán los que estén disponibles en el sitio de internet Kaggle.com.
* La información estadística para comparar se modelará con una regresión lineal, de esta forma poder hacer la comparación.
* La solución estará basada en redes neuronales artificiales, utilizando perceptrón multicapa y en una red profunda con LSTM.

1. **Metodología y herramientas**

## **Metodología**

La metodología utilizada tiene como marco referencial la metodología CRISP (Figura 4), la cual parte por la comprensión del negocio y datos, preparación de los datos, modelado, evaluación e implementación, todo bajo un proceso iterativo hasta encontrar la solución definitiva.

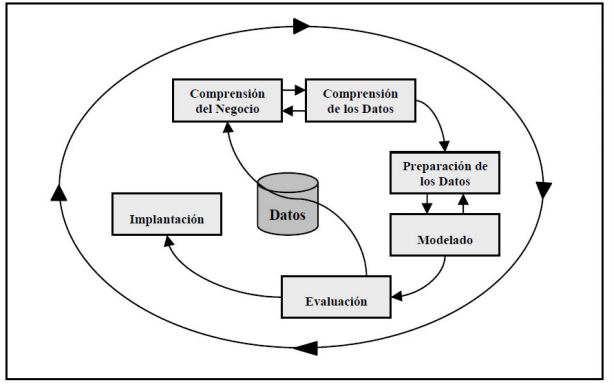


Figura 4: Metodología CRISP [16].

En cuanto al proceso de predicción basado en Machine Learning, tendrá como entrada los datos, existirá un proceso de entrenamiento del modelo creado hasta obtener la predicción esperada. Como se puede visualizar en la Figura 5, los pasos del proceso de predicción son obtener datos, entender datos y su representación, filtrar datos perdidos y fuera de rango (también llamado preprocesamiento), separar datos de prueba y validación, diseñar modelo neuronal, entrenar modelo, analizar resultado y comparar con datos estadísticos existentes [17]. Este proceso será iterativo y se repetirá hasta encontrar la solución esperada.



Figura 5: Pasos de proceso de predicción con ML [17].

## **Herramientas**

Las herramientas que se usarán serán de Data Science disponibles y de código abierto. El primer paquete que será usado es Python para ciencia de datos, el cual se compone por Jupyter Notbook, Python y NumPy. Además, se utilizará la biblioteca Python de análisis de datos: PANDAS, también las herramientas para Machine Learning que provee Scikit-learn y KERAS (biblioteca de aprendizaje profundo para Python). El ambiente necesario para este proyecto de título es un computador MacBook Pro-15’ (Intel Core i7 de 6 núcleos y 2,2 GHz, 16 GB de memoria LPDDR3 y SSD de 256 GB), lo que otorgará un rendimiento que permita ejecutar el modelo respectivo.

## **Plan de trabajo**

Este proyecto de título será trabajado en 600 HH, para esto se le asignarán 4 horas diarias, permitiendo plantear el proyecto en 21 semanas (Figura 6), las que se desglosan en 17 semanas del segundo semestre y 4 semanas adicionales considerando el mes de julio de 2019. De esta forma desarrollar de buena forma el proyecto de título.

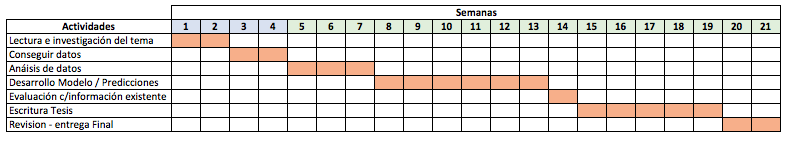


Figura 6: Planificación del trabajo.

1. **Referencias**

[1] Association for Computing Machinery (2018). Fathers of the Deep Learning Revolution Receive ACM A.M. Turing Award. Disponible: <https://awards.acm.org/about/2018-turing>.

[2] [J. Clerc](https://hal.archives-ouvertes.fr/search/index/q/*/authIdHal_s/georges-kariniotakis), [J. Olmedo](https://hal.archives-ouvertes.fr/search/index/q/*/authFullName_s/Georges+Stavrakakis), J. Peralta, M. Saavedra, E. Sauma, I. Urzúa, A. Hernando. “Energías renovables en Chile. Hacia una inserción eficiente de la matriz eléctrica”, Centro de Estudios Públicos (CEP), Chile, 2017.

[3] L. Fernández, “Modelos avanzados para la predicción a corto plazo de la producción eléctrica de parques eólicos”, Tesis Doctoral, Universidad de la Rioja, Logroño, España, 2007.

[4] Ministerio de Energía de Chile (2016). Política Energética de Chile. Disponible: <http://www.energia2050.cl/>.

[5] Ministerio de Energía de Chile (2016). El desarrollo de la energía. Disponible: <https://www.aprendeconenergia.cl/el-desarrollo-de-la-energia-eolica-en-chile/>.

[6] Global Wind Energy Council (2017). Installed Wind Capacity Worlwide. Disponible: <https://gwec.net/publications/global-wind-report-2/>.

[7] E. Lorenz. The Essence of Chaos. University of Washington Press. 1995.

[8] CORFO (2013). Guía de desarrollo de pequeños y medianos proyectos de Energía Eólica. Disponible: <http://dataset.cne.cl/Energia_Abierta/Estudios/CER/Gu%C3%ADa%20eolica%20desarrollo%20de%20proyectos%20peque%C3%B1os%20y%20medianos.pdf>.

[9] J. Patterson y A. Gibson. Deep Learning: A Practitioner's Approach. O’Reilly. 2017.

[10] A. Burkov. The Hundred-Page Machine Learning. Andriy Burkov. 2019.

[11] [C. González](https://hal.archives-ouvertes.fr/search/index/q/*/authIdHal_s/georges-kariniotakis), [D. López](https://hal.archives-ouvertes.fr/search/index/q/*/authFullName_s/Georges+Stavrakakis) (2013). Energía eólica y su intermitencia ¿mayor predicción del viento o almacenamiento de energía?. Disponible: <http://latinoamericarenovable.com/2013/07/27/energia-eolica-y-su-intermitencia-mayor-prediccion-del-viento-o-almacenamiento-de-energia/>.

[12] [H. Mackenzie](https://hal.archives-ouvertes.fr/search/index/q/*/authIdHal_s/georges-kariniotakis), [J. Dyson](https://hal.archives-ouvertes.fr/search/index/q/*/authFullName_s/Georges+Stavrakakis), “Short-Term Forecasting of Wind Power Plant Generation for System Stability and Provision of Ancillary Service”, HARD Software, Australia, 2017.

[13] S. Ortega, “Predicción de extremos de viento en parque eólicos mediante técnicas de machine learning”, Tesis de Máster, Universidad de Alcalá, Madrid, España, 2016.

[14] [G. Kariniotakis](https://hal.archives-ouvertes.fr/search/index/q/*/authIdHal_s/georges-kariniotakis), [G. Stavrakakis](https://hal.archives-ouvertes.fr/search/index/q/*/authFullName_s/Georges+Stavrakakis), [E. Nogaret](https://hal.archives-ouvertes.fr/search/index/q/*/authFullName_s/Eric+Nogaret), “Wind power forecasting using advanced neural networks models”, IEEE Transaction on Energy Conversion, Vol.11, no.4, 762-767, 1996.

[15] I. Erbetta, “Predicción de la potencia para la operación de parques eólicos”, Tesis de Ingeniería Civil Electricista, Universidad de Chile, Santiago, Chile, 2010.

[16] [P. Chapman](https://hal.archives-ouvertes.fr/search/index/q/*/authIdHal_s/georges-kariniotakis), [J. Clinton](https://hal.archives-ouvertes.fr/search/index/q/*/authFullName_s/Georges+Stavrakakis), R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, R. Wirth. “Step-by-step data mining guide”, CRISP-DM Consortium (NCR, DaimlerChrysler, SPSS & Bank Groep B.V.), USA, 2000.

[17] “Curso Redes Neuronales”. Departamento de Ingeniería Informática, Universidad de Santiago de Chile. Segundo semestre 2018.