# INFORME FINAL

# CONCURSO DE INVESTIGACIÓN PARA PREGRADO VERANO 2024

**Este documento deberá ser completado por el/la estudiante, firmado por quienes corresponda y enviado a más tardar el día lunes 01 de abril de 2024, a través del** [**Formulario envío Informe Final IPRE Verano 2024**](https://airtable.com/apppuwyOprhyPOMv1/shr1oV32wWbdTTxNS)

Una vez que sea recibido y validado por la Direccion de Investigación de la Vicerrectoría de Investigación (VRI), se emitirá el certificado de participación y se efectuará el pago del beneficio económico (durante mayo 2024)

**ANTECEDENTES PERSONALES DEL/DE LA ESTUDIANTE**

|  |  |
| --- | --- |
| Nombre completo | Bernardita Paz Rosas Riveros |
| RUN | 21.031.517-9 |
| Unidad Académica | Ingeniería |

**DATOS GENERALES DE LA INVESTIGACIÓN**

|  |  |
| --- | --- |
| Nombre profesor/a guía | Elwin van’t Wout |
| Unidad académica profesor/a guía | Ingeniería matemática |
| Título de Proyecto de investigación | AI supported forecasts of heat waves along the Pacific coast of the Americas |
| Fecha de inicio actividades | 12/12/2023 |
| Fecha de término actividades | 3/02/2024 |
| Horas de trabajo totales | 80 |
| Formato de trabajo | Presencial\_\_\_ Teletrabajo\_\_\_ Híbrido\_X\_\_ |
| Lugar/es en que se desarrollaron las actividades | Campus San Joaquín |
| ¿Pudiste realizar todas las actividades comprometidas en la postulación? | Sí\_\_ No\_X\_ |
| **Actividades y objetivos:** nombra los principales resultados de las actividades realizadas, considerando si los objetivos planteados en la postulación fueron alcanzados. En caso de no haber sido logrados, indica por qué (**máx. 3/4 página,** letra Calibri, tamaño 11). | |
| Las actividades y objetivos realizados variaron de lo presentado inicialmente ya que mi trabajo estuvo enfocado, por sobre todo, en el estudio de la utilización de índices (ENSO, PDO, DMI, entre otros) para la predicción de olas de calor. Para ello, la investigación se centró en la estación de Quinta Normal a una escala anual.  Por un lado, se analizó el nivel de correlación que cada uno de los índices (ENSO, PDO, DMI, AAO y SAM) tiene con cada una de las etiquetas a predecir (número, temperatura, intensidad y duración de olas de calor, entre otras). Como resultado, se evidenció una correlación relativamente pequeña de la mayoría de los predictores con respecto al número de olas de calor (ENSO: 0.01, PDO: -0.11, DMI: 0.4, AAO:0.25, SAM:0.12). Dado lo anterior, se decidió ahondar en porqué el índice ENSO muestra una correlación tan pequeña cuando visualmente se veía que esta era mayor. De esta manera, se intentó otro acercamiento en el que se le realizó un ‘detrending’ al número de olas de calor (y otras etiquetas) para que así los datos tuviesen ciclos que se correlacionasen de mejor manera con índices predictores al remover tendencias debidas al aumento de temperatura. Lo anterior resultó en un aumento de la correlación de ciertos índices como el ENSO, aunque disminuyendo algunos otros como AAO o SAM (ENSO: 0.31, PDO: 0.33, DMI: 0.21, AAO:0.07, SAM:0.0). Un tutorial para realizar el ‘detrending’ de los datos fue realizado en el repositorio del proyecto. Se logró el objetivo de reconocer aquellos predictores que tienen una mayor correlación con la incidencia de indicadores de olas de calor (considerando o no tendencias).  Por otro lado, se entrenó a modelos de predicción de la librería ‘sickit-learn’ de Python. Estos correspondieron a un modelo de regresión linear (LR), un modelo de red neuronal (MLP) y un modelo de red neuronal con inputs ajustados (MLPScaled). A partir de sus resultados, se concluye que hay varios aspectos fundamentales para mejorar el rendimiento de los modelos. En primer lugar, para garantizar que el entrenamiento disminuya su sesgo, es necesario entrenar el modelo con datos aleatorios para aumentar la variabilidad del espacio de entrenamiento. En segundo lugar, resulta importante añadir algún predictor que aumente constantemente (por ejemplo el año), debido al aumento de la temperatura, y por ello, de las olas de calor, en los últimos 20 años. En tercer lugar, es necesario tener en consideración el sobre entrenamiento que experimentan las redes neuronales y tomar medidas de precaución como la simplificación de su arquitectura. De esta manera, el mejor rendimiento que se obtuvo fue con MLP (utilizando como predictores al año, ENSO, PDO, DMI y AAO) y con una elección al azar de los datos, obteniendo un error promedio de 1,6 olas de calor. Cumpliendo el objetivo de implementar los primeros modelos del proyecto para mejorarlos. | |
| **Metodología:** describe la metodología utilizada mencionando si se cumplieron los protocolos en cuanto a ética y seguridad en investigación, en caso que corresponda (**máx. 1/2 página,** letra Calibri, tamaño 11). | |
| En primer lugar, se recopilaron y procesaron valores anuales para todos los índices a estudiar, siendo estos: ENSO, PDO, DMI, AAO y SAM, desde 1981 hasta 2022 (SAM hasta 2015). Estos fueron obtenidos a través de la Oficina Nacional de Administración Oceánica y Atmosférica (NOAA) de Estados Unidos. Adicionalmente, se obtuvo información acerca de los indicadores relacionados a la ocurrencia de olas de calor (número, temperatura, duración e intensidad promedio) desde 1981 hasta 2022 y se crearon nuevos indicadores como número de olas de calor multiplicado por temperatura media, entre otros.  En segundo lugar, se calcularon las correlaciones (spearman y regular) existentes entre los predictores y los indicadores a predecir. A partir de ello, se realizó un ‘detrending’ de los datos, con el objetivo de remover la tendencia positiva que los datos presentan desde el año 2000 debido al aumento de las temperaturas. Se volvieron a calcular las correlaciones tras ello.  En tercer lugar, se entrenaron modelos de predicción de la librería ‘sickit-learn’ de Python, estos fueron; un modelo de regresión linear (LR), un modelo de red neuronal (MLP) y un modelo de red neuronal con inputs ajustados (MLPScaled). Adicionalmente, se analizó el rendimiento de tales modelos para mejorar sus resultados a través de cambios a estos o su entrenamiento.  En cuanto a implicancias éticas, los datos utilizados fueron tomados desde fuentes internacionales que no cuentan con datos sensibles. | |
| **Plan de trabajo:** Indica brevemente cómo se ejecutó el plan de trabajo, y si éste se realizó de acuerdo a lo planificado (**máx. 1/2 página,** letra Calibri, tamaño 11). | |
| En un comienzo, se realizó una reunión con el profesor encargado en la cuál se definió que la investigación tendría el objetivo de ‘Diseñar un algoritmo predictivo de la frecuencia e intensidad anual de olas de calor en Chile hasta el año 2100.’ Para ello se estableció que; se buscarían predictores de olas de calor en la literatura (tales como ENSO), se crearían ‘features’ para cuantificar tales predictores de manera anual, se diseñarían algoritmos de aprendizaje automatizado a través de Python, se realizarían proyecciones a futuro de olas de calor y, por último, se haría una predicción de la frecuencia e intensidad de olas de calor en Chile hasta 2100.  En cuanto al seguimiento el plan, este fue bastante útil, aunque finalmente, no se realizaron predicciones futuras al existir un mayor interés en el mejoramiento de los algoritmos a partir de datos históricos. En definitiva, gran parte del proyecto de investigación surgió a partir de interrogantes acerca de cómo mejorar y visualizar las relaciones existentes entre las variables estudiadas para ser un aporte al macro-proyecto. | |

Firma estudiante: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Firma profesor/a guía: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_



Nombre director/a de Investigación\*: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Firma director/a de Investigación: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\*Corresponde al/a la director/a de investigación de la unidad académica del/de la profesor/a guía.