



CC57- MACHINE LEARNING

Trabajo Parcial / Final 2024

SECCIONES: TODAS

PROFESORES:	Fernandez Vasquez, Richard Fernando Reyes Silva, Patricia Daniela
SECCIONES:	Todas
FECHA DE EVALUACIÓN:	Por hitos: Hito1 (semana 7), Hito2 (semana 13) e Hito 3 (semana 15)

1. Objetivo

El presente documento define el trabajo Parcial/Final y la rúbrica que permite evaluar el logro del curso **CC57 – MACHINE LEARNING**.

2. Logro del curso

PARA EVALUACION DE TRABAJO FINAL

Competencias Generales: Pensamiento Innovador

Nivel de logro: 2

El estudiante adquiere la capacidad de desarrollar y llevar a cabo la experimentación adecuada, analizar e interpretar datos, y usar el juicio de ingeniería para sacar conclusiones.

PARA EVALUACION DE DESEMPEÑO

Competencias Específicas: ABET 6 - Técnicas en computación

Nivel de logro: 1

APLICAR LA TEORÍA DE CIENCIAS DE LA COMPUTACIÓN Y LOS FUNDAMENTOS DEL DESARROLLO DE SOFTWARE PARA PRODUCIR SOLUCIONES BASADAS EN COMPUTACIÓN.

3. Enunciado

El Trabajo Parcial/Final consiste en aplicar técnicas de Machine Learning desarrolladas durante el presente ciclo de estudios, proponiendo un Modelo de Machine Learning, acorde con las competencias generales y específicas del curso.

Como un proyecto de analítica, el modelo de ML creado deberá proporcionar conocimiento a partir de los datos, y con ello, poder responder a las preguntas que nuestro cliente, **Perú C-Inversiones** solicita.



CC57- MACHINE LEARNING

Trabajo Parcial / Final 2024

SECCIONES: TODAS



Perú C-Inversiones es una empresa administradora de capitales privados peruanos que desea brindar a sus clientes un producto de inversión compuesto por un conjunto de criptomonedas de baja capitalización y, por consiguiente, de mayor riesgo.

La empresa actualmente ofrece a sus clientes un producto de inversión que incluye solo a Bitcoin y Ethereum como cripto activos, sin embargo, conociendo que el cuarto evento de halving de Bitcoin tendrá lugar en abril del 2024, desea contar con un modelo de inteligencia artificial que le proponga un conjunto de cripto activos de baja capitalización y que posean una alta probabilidad de incrementar notablemente su valorización actual en los siguientes meses.

Como conocedor del mercado de criptomonedas, es de su interés invertir solo en las siguientes narrativas:

- Inteligencia Artificial
- Videojuegos
- Activos del mundo real (RWA o real world assets)
- Memes

¡Ayudemos a **Perú C-Inversiones**!

4. Estructura del Informe

El informe a presentar debe estar escrito con la siguiente estructura:

Situación de contexto real, Adquisición de los Datos, Ingeniería de Características, Propuesta Técnica, Experimentos, Validación de resultados y pruebas, Conclusiones y Referencias bibliográficas.

1. **Situación de contexto real.** Redactar la descripción y fundamentación de la situación real, utilizando como fuentes los informes correspondientes a las Practicas Calificadas 1 y 2 (PC1/PC2).
2. **Adquisición y Preprocesamientos de los Datos.**
 - Cada grupo obtendrá los datos según requerimiento del Científico de Datos, a partir de distintas fuentes/origen propuestas (ver detalle en Anexo 1).
 - En esta sección, se debe señalar y describir el conjunto de datos que se le solicita adquirir, así como las técnicas/procesos utilizados para su obtención.
 - Es de **carácter obligatorio** documentar las fuentes y estructura de los datos adquiridos.
 - Cada grupo debe garantizar el debido **preprocesamiento y calidad de los datos** obtenidos, para lo cual, adjuntará el respectivo código (y enlace al repositorio a GitHub) que evidencia la adquisición y preprocesamiento realizado (recuerde que los datos formarán parte de un repositorio común, del cual todos los grupos se servirán



CC57- MACHINE LEARNING

Trabajo Parcial / Final 2024

SECCIONES: TODAS

para la creación de sus modelos). Será importante, realizar un conjunto de visualizaciones a fin de encontrar algún patrón interesante.

3. **Propuesta.** Como equipo, y siguiendo la metodología de trabajo CRISP-DM considerar:
 - Describir las tareas a llevar a cabo en cada una de las fases de este proyecto (incluir cronograma de trabajo).
 - Describir las técnicas o algoritmos de Machine Learning que la propuesta podrá seleccionar y posteriormente entrenar a partir de los datos a fin de obtener el conocimiento esperado por el cliente.
4. **Ingeniería de Características.**
 - La selección de las características dependerá de las técnicas de ML que cada grupo propondrá como solución dentro de su propuesta.
 - Los datos podrán obtenerse desde el repositorio común de datos.
 - Solo en el caso que la propuesta lo requiera y no se disponga de datos, el grupo deberá adquirir los datos necesarios no contemplados en el repositorio común.
5. **Experimentos.** La propuesta deberá contemplar un conjunto de experimentos que incluya la configuración del algoritmo para su entrenamiento y/o cambios necesarios para su ejecución.
6. **Validación de resultados y pruebas.** Describir las entradas y salidas, interpretación de resultados y pruebas del modelo de ML creado.
7. **Comunicación.** Los resultados obtenidos serán comunicados al cliente a través de una interfase de usuario básica amigable. Sean los resultados positivos o no, deberán ser comunicados.
8. **Conclusiones.** En un párrafo redactar las conclusiones del trabajo, especificando la(s) técnica(s) utilizadas y, la comunicación de los resultados, y de ser el caso, el trabajo a futuro.
9. **Referencias bibliográficas.** Considere un mínimo de tres trabajos relacionados que deberán ser citados como referencia bibliográfica en su informe.

El informe debe contener como mínimo 8 y máximo 15 páginas.

5. Acerca del grupo de trabajo

El trabajo se deberá desarrollar en grupo de 3-4, según sea el caso.

6. Lenguaje de programación

El desarrollo de la aplicación debe estar implementado en Python salvo alguna restricción sustentada puede usar otro lenguaje. El modelo de ML será desplegado a partir de interfaz gráfica (GUI) básica, implementada en cualquier lenguaje de programación.

7. Exposición

- La exposición es parte de la evaluación del trabajo final que se realizará en la última sesión de la semana 15.
- Con Vestimenta formal.
- Sera evaluada la claridad y organización de su presentación (ayuda audiovisual).
- 15 minutos de exposición por cada grupo.
- Se realizarán preguntas a los integrantes del grupo acerca del trabajo y deberán detallar el



CC57- MACHINE LEARNING

Trabajo Parcial / Final 2024

SECCIONES: TODAS

contenido.

8. Evaluación del Trabajo Parcial/Final

El trabajo se ha dividido en **3 hitos**:

PRIMER HITO (TRABAJO PARCIAL):

- **Escribir el informe** de acuerdo al siguiente contenido:
 1. **Situación de contexto real.** Redactar la descripción y fundamentación de la situación real, utilizando como fuentes los informes correspondientes a las Practicas Calificadas 1 y 2 (PC1/PC2).
 2. **Adquisición y Preprocesamientos de los Datos.**
 - Cada grupo obtendrá los datos según requerimiento del Científico de Datos, a partir de distintas fuentes/origen propuestas (ver detalle en Anexo 1).
 - En esta sección, se debe señalar y describir el conjunto de datos que se le solicita adquirir, así como las técnicas/procesos utilizados para su obtención.
 - Es de **carácter obligatorio** documentar las fuentes y estructura de los datos adquiridos.
 - Cada grupo debe garantizar el debido **preprocesamiento y calidad de los datos** obtenidos, para lo cual, adjuntará el respectivo código (y enlace al repositorio a GitHub) que evidencia la adquisición y preprocesamiento realizado (recuerde que los datos formarán parte de un repositorio común, del cual todos los grupos se servirán para la creación de sus modelos). Será importante, realizar un conjunto de visualizaciones a fin de encontrar algún patrón interesante.
 3. **Propuesta.** Como equipo, y siguiendo la metodología de trabajo CRISP-DM considerar:
 - Describir las tareas a llevar a cabo en cada una de las fases de este proyecto (incluir cronograma de trabajo).
 - Describir las técnicas o algoritmos de Machine Learning que la propuesta podrá seleccionar y posteriormente entrenar a partir de los datos a fin de obtener el conocimiento esperado por el cliente.
- **Entrega en Aula virtual:**

Crear un archivo con el nombre TP_XXX_YYY_ZZZ.zip, donde los caracteres X, Y y Z corresponden al primer apellido de los integrantes del grupo.

 - Entregar archivo en el Aula Virtual en Fecha: **Semana 7**
 - Contenido del archivo:
 - Informe en formato **TP_XXX_YYY_ZZZ.pdf**
 - Archivo texto conteniendo nombre / ruta del archivo de datos adquirido
Data_XXX_YYY_ZZZ.txt



CC57- MACHINE LEARNING

Trabajo Parcial / Final 2024

SECCIONES: TODAS

- Puntaje asignado: 0 a 20 puntos

SEGUNDO HITO:

1. **Situación de contexto real.** Redactar la descripción y fundamentación de la situación real, utilizando como fuentes los informes correspondientes a las Practicas Calificadas 1 y 2 (PC1/PC2).
2. **Adquisición y Preprocesamientos de los Datos.**
 - Cada grupo obtendrá los datos según requerimiento del Científico de Datos, a partir de distintas fuentes/origen propuestas (ver detalle en Anexo 1).
 - En esta sección, se debe señalar y describir el conjunto de datos que se le solicita adquirir, así como las técnicas/procesos utilizados para su obtención.
 - Es de **carácter obligatorio** documentar las fuentes y estructura de los datos adquiridos.
 - Cada grupo debe garantizar el debido **preprocesamiento y calidad de los datos** obtenidos, para lo cual, adjuntará el respectivo código (y enlace al repositorio a GitHub) que evidencia la adquisición y preprocesamiento realizado (recuerde que los datos formarán parte de un repositorio común, del cual todos los grupos se servirán para la creación de sus modelos). Será importante, realizar un conjunto de visualizaciones a fin de encontrar algún patrón interesante.
3. **Propuesta.** Como equipo, y siguiendo la metodología de trabajo CRISP-DM considerar:
 - Describir las tareas a llevar a cabo en cada una de las fases de este proyecto (incluir cronograma de trabajo).
 - Describir las técnicas o algoritmos de Machine Learning que la propuesta podrá seleccionar y posteriormente entrenar a partir de los datos a fin de obtener el conocimiento esperado por el cliente.
4. **Ingeniería de Características.**
 - La selección de las características dependerá de las técnicas de ML que cada grupo propondrá como solución dentro de su propuesta.
 - Los datos podrán obtenerse desde el repositorio común de datos.
 - Solo en el caso que la propuesta lo requiera y no se disponga de datos, el grupo deberá adquirir los datos necesarios no contemplados en el repositorio común.
5. **Experimentos.** La propuesta deberá contemplar un conjunto de experimentos que incluya la configuración del algoritmo para su entrenamiento y/o cambios necesarios para su ejecución.

- **Entrega en Aula virtual:**

Crear un archivo con el nombre TF_XXX_YYY_ZZZ, donde los caracteres X, Y y Z corresponden al primer apellido de los integrantes del grupo.

- Fecha: **Semana 13**
- Rubrica de evaluación: **Rubrica de Trabajo Final**
- Puntaje asignado: 0 a 5 puntos

TERCER HITO (TRABAJO FINAL):



CC57- MACHINE LEARNING

Trabajo Parcial / Final 2024

SECCIONES: TODAS

- **El informe debe estar escrito de acuerdo al siguiente contenido:**
 1. **Situación de contexto real.** Redactar la descripción y fundamentación de la situación real, utilizando como fuentes los informes correspondientes a las Practicas Calificadas 1 y 2 (PC1/PC2).
 2. **Adquisición y Preprocesamientos de los Datos.**
 - Cada grupo obtendrá los datos según requerimiento del Científico de Datos, a partir de distintas fuentes/origen propuestas (ver detalle en Anexo 1).
 - En esta sección, se debe señalar y describir el conjunto de datos que se le solicita adquirir, así como las técnicas/procesos utilizados para su obtención.
 - Es de **carácter obligatorio** documentar las fuentes y estructura de los datos adquiridos.
 - Cada grupo debe garantizar el debido **preprocesamiento y calidad de los datos** obtenidos, para lo cual, adjuntará el respectivo código (y enlace al repositorio a GitHub) que evidencia la adquisición y preprocesamiento realizado (recuerde que los datos formarán parte de un repositorio común, del cual todos los grupos se servirán para la creación de sus modelos). Será importante, realizar un conjunto de visualizaciones a fin de encontrar algún patrón interesante.
 3. **Propuesta.** Como equipo, y siguiendo la metodología de trabajo CRISP-DM considerar:
 - Describir las tareas a llevar a cabo en cada una de las fases de este proyecto (incluir cronograma de trabajo).
 - Describir las técnicas o algoritmos de Machine Learning que la propuesta podrá seleccionar y posteriormente entrenar a partir de los datos a fin de obtener el conocimiento esperado por el cliente.
 4. **Ingeniería de Características.**
 - La selección de las características dependerá de las técnicas de ML que cada grupo propondrá como solución dentro de su propuesta.
 - Los datos podrán obtenerse desde el repositorio común de datos.
 - Solo en el caso que la propuesta lo requiera y no se disponga de datos, el grupo deberá adquirir los datos necesarios no contemplados en el repositorio común.
 5. **Experimentos.** La propuesta deberá contemplar un conjunto de experimentos que incluya la configuración del algoritmo para su entrenamiento y/o cambios necesarios para su ejecución.
 6. **Validación de resultados y pruebas.** Describir las entradas y salidas, interpretación de resultados y pruebas del modelo de ML creado.
 7. **Comunicación.** Los resultados obtenidos serán comunicados al cliente a través de una interfase de usuario básica amigable. Sean los resultados positivos o no, deberán ser comunicados.
 8. **Conclusiones.** En un párrafo redactar las conclusiones del trabajo, especificando la(s) técnica(s) utilizadas y, la comunicación de los resultados, y de ser el caso, el trabajo a futuro.
 9. **Referencias bibliográficas.** Considere un mínimo de tres trabajos relacionados que deberán ser citados como referencia bibliográfica en su informe.
- **Entrega del trabajo final en el Aula virtual:**



CC57- MACHINE LEARNING

Trabajo Parcial / Final 2024

SECCIONES: TODAS

Crear una carpeta con el nombre TF_XXX_YYY_ZZZ, donde los caracteres X, Y y Z corresponden al primer apellido de los integrantes del grupo. La estructura de la carpeta es el siguiente:

- ✓ Sub carpeta de **code**
- ✓ Sub carpeta **data**
- ✓ Link de video grabado de la exposición 12 minutos (cada uno 4 minutos) en un archivo de texto.
- ✓ TF_Informe_XXXXX_YYYYY.docx

Empaquetar o comprimir la carpeta TF_XXX_YYY_ZZZ.zip entregar en Blackboard.

- Fecha: **Semana 15 (EXPOSICIÓN DURANTE LA SEMANA)**
- Rubrica de evaluación: **Rúbrica de trabajo final.**
- Puntaje asignado: 0 a 15 puntos.

Setiembre 2024