

Pautas para tesis

Maestría en Ciencia de Datos 2024

El presente documento tiene como objetivo presentar las pautas para la tesis. La misma es obligatoria para la obtención del título de Maestría.

En términos generales, la misma deberá ser de carácter aplicado, enfocándose en la aplicación de técnicas de Data Science¹ para resolver un caso de negocio específico.

Contenido

A continuación, se detalla una estructura deseable para dicho trabajo:

1. Portada

Es la primera página y cuenta con la información básica:

- A) Logo institucional de Universidad de Montevideo y UMPE
- B) Identificación del Programa en el marco del cual realiza la tesis (Maestría en Ciencia de Datos)
- C) Título del trabajo: debe reflejar el alcance de la investigación, siendo claro, conciso y específico.
- D) Autor/es: nombres y apellidos completos de los integrantes del equipo, ordenados alfabéticamente por los apellidos de los autores.
- E) Tutor: se indican los nombres y apellidos completos del profesor asignado.
- F) Lugar y fecha.

2. Resumen y palabras clave

El resumen debe caracterizar al trabajo en un rango de 150 a 300 palabras. El objetivo del resumen es destacar la pertinencia de la investigación y permite decidir al lector si el documento es de su interés.

Las palabras clave son términos representativos que identifican el contenido del trabajo. Se especifican a continuación del resumen.

¹ Se pretende que el alumno aplique técnicas de Inteligencia Artificial en alguna de sus diversas áreas (Machine Learning, Deep Learning, NLP, LLM, Computer Vision, etc.).

3. Introducción

La introducción es la presentación general, clara, y breve del contenido del trabajo, conteniendo:

- Presentación del problema de negocio a resolver, incluyendo su relevancia, motivación y antecedentes
- Objetivos generales e hipótesis
- Metodología utilizada para lograr los objetivos propuestos

4. Metodología

Se trata de una descripción detallada de las técnicas de Inteligencia Artificial (Machine Learning, Deep Learning, NLP, LLM, Computer Vision, etc.) y herramientas utilizadas. Las mismas deben estar justificadas.

5. Análisis exploratorio de datos (EDA).

Consiste en la presentación de las principales conclusiones de una exploración inicial de los datos. Se detallan los datos disponibles y se realiza una descripción estadística de los mismos.

Se sugiere la utilización de gráficos y tablas en esta sección.

6. Conceptos Teóricos - Estado del Arte

Revisión de la literatura relacionada con el problema de negocio específico que se plantea resolver, así como también los fundamentos teóricos de las técnicas utilizadas.

7. Experimentos

Descripción de los experimentos realizados y los modelos estadísticos aplicados, detallando claramente los resultados obtenidos en términos de métrica útiles para selección de modelos.

8. Conclusiones

Discusión sobre la eficacia de las soluciones propuestas y elaboración de recomendaciones para futuras investigaciones o aplicaciones en el negocio.

Puntos a valorar para la evaluación

1. Técnicas utilizadas

- Correcta implementación de las técnicas y relevancia.
- Innovación y creatividad en la aplicación de estas técnicas.
- Precisión y rendimiento de los modelos desarrollados.

2. Fundamentación teórica

- Rigor científico en la presentación de conceptos teóricos.
- Justificación teórica de las decisiones metodológicas.

3. Resolución de un problema de negocio relevante

- Relevancia y claridad del problema de negocio abordado.
- Impacto potencial de la solución en el negocio.

4. Aspectos generales

- Escritura académica
- Correcto manejo de citas, bibliografía, etc.
- Correcto manejo de tablas y gráficos

Se sugiere el uso de GitHub (o similares) como repositorio de trabajo, disponibilizando el mismo al momento de la entrega para asegurar la reproducibilidad de los resultados en caso de ser necesario.

Supervisión y seguimiento

Cada grupo de tesis contará con el apoyo de un Tutor Académico que oficiará de supervisor y se encargará de orientar y coordinar las actividades del grupo. Se coordinarán reuniones para asegurar el seguimiento adecuado del progreso de la tesis y orientar a los alumnos durante el desarrollo del proyecto.

El equipo de trabajo puede optar por trabajar en Python y R, siendo estos los únicos lenguajes de programación aceptados para el análisis.

Organización y plazos

La tesis se realizará en grupos de hasta tres integrantes. Se plantea un plazo de trabajo de 5 meses, en los que el estudiante tendrá como prioridad la realización de la tesis.

Se espera un rol activo del Tutor en materia de seguimiento semanal, así como una alta dedicación del equipo de estudiantes, para asegurar la realización en los plazos estipulados.

La elección del tema es responsabilidad del equipo y el mismo puede ser ajustado en base a sugerencias del Tutor y/o la Coordinación de la Maestría. Podrán utilizarse datos propios (empresariales, justificando debidamente su origen y reservándose la Universidad el derecho a solicitar la firma de un contrato que habilite al uso de los mismos) o utilizar un dataset disponible públicamente (Kaggle, Datos abiertos Municipales o Nacionales, Gapminder u otra fuente que el alumno considere).²

La entrega consiste en la presentación del documento de trabajo (formato PDF) y el acceso al repositorio que da sustento a las conclusiones allí reflejadas.

La fecha de entrega y defensa se encuentra estipulada en el calendario académico para el mes de diciembre. Quienes no logren finalizar sus trabajos para dicha fecha, podrán hacerlo nuevamente en abril del próximo año. Excepcionalmente se podrá extender la entrega hasta julio, contando con aprobación y justificación por parte del tutor.

² La sugerencia es que, en cuanto sea posible, se trabaje con datos propios para garantizar que el componente innovador del trabajo. En caso de trabajar con datos públicos, se recomienda una selección apropiada del mismo para evitar caer en ejemplos altamente reutilizados.