

GENERACIÓN DE DATOS SINTÉTICOS CON AUTOENCODERS  
VARIACIONALES Y SU IMPACTO EN LA SELECCIÓN DE  
CARACTERÍSTICAS MEDIANTE ALGORITMOS GENÉTICOS

Tesis para optar por el título de Maestría en  
Minería de Datos

Lic. Claudio Sebastián Castillo

Dirigida por:  
Dr. Matías Gerard - Director  
Dr. Leandro Vignolo - Co-Director

Paraná, Entre Ríos, Argentina  
2024

*I, AUTHORMNAME confirm that the work presented in this thesis is my own.  
Where information has been derived from other sources, I confirm that this has  
been indicated in the thesis.*

# **Abstract**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam et turpis gravida, lacinia ante sit amet, sollicitudin erat. Aliquam efficitur vehicula leo sed condimentum. Phasellus lobortis eros vitae rutrum egestas. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Donec at urna imperdiet, vulputate orci eu, sollicitudin leo. Donec nec dui sagittis, malesuada erat eget, vulputate tellus. Nam ullamcorper efficitur iaculis. Mauris eu vehicula nibh. In lectus turpis, tempor at felis a, egestas fermentum massa.

# Acknowledgements

Interdum et malesuada fames ac ante ipsum primis in faucibus. Aliquam congue fermentum ante, semper porta nisl consectetur ut. Duis ornare sit amet dui ac faucibus. Phasellus ullamcorper leo vitae arcu ultricies cursus. Duis tristique lacus eget metus bibendum, at dapibus ante malesuada. In dictum nulla nec porta varius. Fusce et elit eget sapien fringilla maximus in sit amet dui.

Mauris eget blandit nisi, faucibus imperdiet odio. Suspendisse blandit dolor sed tellus venenatis, venenatis fringilla turpis pretium. Donec pharetra arcu vitae euismod tincidunt. Morbi ut turpis volutpat, ultrices felis non, finibus justo. Proin convallis accumsan sem ac vulputate. Sed rhoncus ipsum eu urna placerat, sed rhoncus erat facilisis. Praesent vitae vestibulum dui. Proin interdum tellus ac velit varius, sed finibus turpis placerat.

# Índice general

<b>Abstract</b>	I
<b>Acknowledgements</b>	II
<b>1. Introduction, with a citation</b>	1
1.1. Background . . . . .	1
1.2. The middle bit . . . . .	2
1.2.1. Subsection of the middle bit . . . . .	2
1.3. Summary of chapters . . . . .	2
<b>2. Literature review, with maths</b>	4
2.1. Introduction . . . . .	4
2.2. The middle . . . . .	4
2.3. A complicated math equation . . . . .	5
2.4. Conclusion . . . . .	5
<b>3. Algoritmos clásicos y sus resultados</b>	6
3.1. Datos elegidos en nuestro estudio . . . . .	7
3.2. Modelos Elegidos . . . . .	8
3.2.1. Subsection 2 . . . . .	8
3.3. Entorno de Experimentación . . . . .	9
3.4. Results . . . . .	9
3.5. Discussion . . . . .	9
3.6. Conclusion . . . . .	9
<b>4. Research containing a figure</b>	10
4.1. Introduction . . . . .	10
4.2. Method . . . . .	10
4.2.1. Subsection 1 . . . . .	10
4.2.2. Subsection 2 . . . . .	11
4.3. Results . . . . .	11
4.4. Discussion . . . . .	11
4.5. Conclusion . . . . .	11
<b>5. Research containing a table</b>	14
5.1. Introduction . . . . .	14
5.2. Method . . . . .	14

5.2.1. Subsection 1 . . . . .	15
5.2.2. Subsection 2 . . . . .	15
5.3. Results . . . . .	15
5.4. Discussion . . . . .	16
5.5. Conclusion . . . . .	17
<b>6. Final research study</b>	<b>18</b>
6.1. Introduction . . . . .	18
6.2. Method . . . . .	18
6.2.1. Subsection 1 . . . . .	18
6.2.2. Subsection 2 . . . . .	19
6.3. Results . . . . .	19
6.4. Discussion . . . . .	19
6.5. Conclusion . . . . .	19
<b>7. Conclusion</b>	<b>20</b>
7.1. Thesis summary . . . . .	20
7.2. Future work . . . . .	20
<b>Appendix 1: Some extra stuff</b>	<b>22</b>
<b>Appendix 2: Some more extra stuff</b>	<b>23</b>
<b>References</b>	<b>24</b>

# Capítulo 1

## Introduction, with a citation

### 1.1. Background

This is the introduction. Quisque finibus aliquet cursus. Integer in pellentesque tellus. Duis eu dignissim nulla, a porttitor enim. Quisque vehicula leo non ultrices finibus. Duis vehicula quis sem sit amet sollicitudin. Integer neque est, pharetra et auctor vel, iaculis interdum lectus.

To include a citation to the text, just add the citation key shown in the references.bib file. The style of the citation is determined by the ref\_format.csl file. For example, in The Living Sea you can find pictures of the Calypso ([Cousteau1963?](#)).

In neque mauris, maximus at sapien a, iaculis dignissim justo. Aliquam erat volutpat. Praesent varius risus auctor est ultricies, sit amet consequat nisi laoreet. Suspendisse non est et mauris pharetra sagittis non porta justo. Praesent malesuada metus ut sapien sodales ornare.

## 1.2. The middle bit

This is the middle bit. Phasellus quis ex in ipsum pellentesque lobortis tincidunt sed massa. Nullam euismod sem quis dictum condimentum. Suspendisse risus metus, elementum eu congue quis, viverra ac metus. Donec non lectus at lectus euismod laoreet pharetra semper dui. Donec sed eleifend erat, vel ultrices nibh. Nam scelerisque turpis ac nunc mollis, et rutrum nisl luctus.

Duis faucibus vestibulum elit, sit amet lobortis libero. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Sed at cursus nibh. Sed accumsan imperdiet interdum. Proin id facilisis tortor. Proin posuere a neque nec iaculis. Suspendisse potenti. Nullam hendrerit ante mi, vitae iaculis dui laoreet eu.

Cras eleifend velit diam, eu viverra mi volutpat ut. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec finibus leo nec dui imperdiet, tincidunt ornare orci venenatis. Maecenas placerat efficitur est, eu blandit magna hendrerit eu.

### 1.2.1. SUBSECTION OF THE MIDDLE BIT

This is a subsection of the middle bit. Quisque sit amet tempus arcu, ac suscipit ante. Cras massa elit, pellentesque eget nisl ut, malesuada rutrum risus. Nunc in venenatis mi. Curabitur sit amet suscipit eros, non tincidunt nibh. Phasellus lorem lectus, iaculis non luctus eget, tempus non risus. Suspendisse ut felis mi.

## 1.3. Summary of chapters

This is a brief outline of what went into each chapter, and a section which shows how to reference headers (which are labelled automatically for you). This

chapter, Section 1, shows how to use citations and how to reference section headers. Section 2 shows how use and reference equations. Section ?? shows how to use and reference code. Section 4 shows how to use, reference, and resize pdf and jpg figures. Section 5 shows how to use and reference tables. Section 6 is truly revolutionary (but shows nothing functional). **Appendix 1** shows how to add chapters which are not numbered, and has to be referenced manually, as does **Appendix 2**. See the base `README.md` for how References are handled - leave `*_references.md` alone, and provide it to `pandoc` last.

Proin faucibus nibh sit amet augue blandit varius.

# Capítulo 2

## Literature review, with maths

### 2.1. Introduction

This is the introduction. Duis in neque felis. In hac habitasse platea dictumst. Cras eget rutrum elit. Pellentesque tristique venenatis pellentesque. Cras eu dignissim quam, vel sodales felis. Vestibulum efficitur justo a nibh cursus eleifend. Integer ultrices lorem at nunc efficitur lobortis.

### 2.2. The middle

This is the literature review. Nullam quam odio, volutpat ac ornare quis, vestibulum nec nulla. Aenean nec dapibus in mL/min<sup>-1</sup>. Mathematical formula can be inserted using Latex and can be automatically numbered:

$$f(x) = ax^3 + bx^2 + cx + d \quad (2.1)$$

Nunc eleifend, ex a luctus porttitor, felis ex suscipit tellus, ut sollicitudin sapien purus in libero. Nulla blandit eget urna vel tempus. Praesent fringilla dui

sapien, sit amet egestas leo sollicitudin at.

Later on in the text, you can reference Equation 2.1 and its mind-blowing ramifications. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Sed faucibus pulvinar volutpat. Ut semper fringilla erat non dapibus. Nunc vitae felis eget purus placerat finibus laoreet ut nibh.

### 2.3. A complicated math equation

The following raw text in markdown behind Equation 2.2 shows that you can fall back on L<sup>A</sup>T<sub>E</sub>X if it is more convenient for you. Note that this will only be rendered in `thesis.pdf`

$$\hat{\theta}_g = \operatorname{argmín}_{\theta_g} \left\{ - \sum_{n=1}^N \left( 1 - \mathbb{1}[f(\mathbf{x}^{(n)})] \right) \log f(\mathbf{x}^{(n)} + g(\mathbf{x}^{(n)}; \theta_g)) + \lambda |g(\mathbf{x}^{(n)}; \theta_g)|_2 \right\}, \quad (2.2)$$

### 2.4. Conclusion

This is the conclusion. Donec pulvinar molestie urna eu faucibus. In tristique ut neque vel eleifend. Morbi ut massa vitae diam gravida iaculis. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

- first item in the list
- second item in the list
- third item in the list

# Capítulo 3

## Algoritmos clásicos y sus resultados

En este capítulo revisaremos el desempeño de algoritmos o modelos clásicos en la solución de los problemas de clasificación planteados en los dataset elegidos para nuestra investigación. A tal fin describiremos brevemente la composición de los conjuntos de datos, los algoritmos seleccionados para su tratamiento y los resultados obtenidos para cada uno. Luego, analizando y comparando dichos resultados, elegiremos aquellos con mejor desempeño en las tareas de clasificación considerando no solo su eficacia, sino también su rapidez y consistencia a lo largo de las distintas tareas.

El propósito de esta etapa del trabajo es doble. Por un lado, identificar los modelos más apropiados para servir de *función de fitness* en la implementación de nuestro algoritmo genético. Esto permitirá construir una implementación robusta, que cuenta con una función efectiva y computacionalmente conveniente para evaluar cada solución. Por el otro, disponer de métricas acerca del desempeño

que logran distintas estrategias de clasificación, a partir de las cuales comparar el resultado de nuestras propias soluciones.

### 3.1. Datos elegidos en nuestro estudio

El conjunto de datos elegidos en este trabajo incluye:

1. *Madelon*: conjunto artificial de datos con 500 características, donde el objetivo es un XOR multidimensional con 5 características relevantes y 15 características resultantes de combinaciones lineales de aquellas (i.e. 20 características redundantes). Las otras 480 características fueron generadas aleatoriamente (no tienen poder predictivo). Madelon es un problema de clasificación de dos clases con variables de entrada binarias dispersas. Las dos clases están equilibradas, y los datos se dividen en conjuntos de entrenamiento y prueba. Fue creado para el desafío de Selección de Características NIPS\_2003, y está disponible en el Repositorio UCI. Los datos están divididos en un conjunto de entrenamiento y un conjunto de testeo.
2. *Gisette*: es un dataset creado para trabajar el problema de reconocimiento de dígitos escritos a mano (Isabelle Guyon 2004). Este conjunto de datos forma parte de los cinco conjuntos utilizados en el desafío de selección de características de NIPS 2003. Tiene 13500 observaciones y 5000 atributos. El desafío radica en diferenciar los dígitos ‘4’ y ‘9’, que suelen ser fácilmente confundibles entre sí. Los dígitos han sido normalizados en tamaño y centrados en una imagen fija de 28x28 píxeles. Además, se crearon características de orden superior como productos de estos píxeles para sumergir el problema en un espacio de características de mayor dimensión. También se añadieron características distractoras denominadas “sondas”, que no tienen poder predictivo. El orden de las características y patrones fue aleatorizado.

Los datos están divididos en un conjunto de entrenamiento y un conjunto de testeo.

3. *Leukemia*: El análisis de datos de expresión génica obtenidos de micro-datos de ADN se estudia en Golub (1999) para la clasificación de tipos de cáncer. Construyeron un conjunto de datos con 7129 mediciones de expresión génica en las clases ALL (leucemia linfocítica aguda) y AML (leucemia mielogénica aguda). El problema es distinguir entre estas dos variantes de leucemia (ALL y AML). Los datos se dividen originalmente en dos subconjuntos: un conjunto de entrenamiento y un conjunto de testeo.
4. *GCM*: El conjunto de datos GCM fue compilado en Ramaswamy (2001) y contiene los perfiles de expresión de 198 muestras de tumores que representan 14 clases comunes de cáncer humano<sup>3</sup>. Aquí el enfoque estuvo en 190 muestras de tumores después de excluir 8 muestras de metástasis. Finalmente, cada matriz se estandarizó a una media de 0 y una varianza de 1. El conjunto de datos consta de un total de 190 instancias, con 16063 atributos (biomarcadores) cada una, y distribuidos en 14 clases desequilibradas. Los datos están divididos en un conjunto de entrenamiento y un conjunto de testeo.

### 3.2. Modelos Elegidos

#### 3.2.1. SUBSECTION 2

By running the code in Section ??, we solved AI completely. This is the second part of the methodology. Proin tincidunt odio non sem mollis tristique. Fusce pharetra accumsan volutpat. In nec mauris vel orci rutrum dapibus nec ac nibh. Praesent malesuada sagittis nulla, eget commodo mauris ultricies eget.

Suspendisse iaculis finibus ligula.

### 3.3. Entorno de Experimentación

### 3.4. Results

These are the results. Ut accumsan tempus aliquam. Sed massa ex, egestas non libero id, imperdiet scelerisque augue. Duis rutrum ultrices arcu et ultricies. Proin vel elit eu magna mattis vehicula. Sed ex erat, fringilla vel feugiat ut, fringilla non diam.

### 3.5. Discussion

This is the discussion. Duis ultrices tempor sem vitae convallis. Pellentesque lobortis risus ac nisi varius bibendum. Phasellus volutpat aliquam varius. Mauris vitae neque quis libero volutpat finibus. Nunc diam metus, imperdiet vitae leo sed, varius posuere orci.

### 3.6. Conclusion

This is the conclusion to the chapter. Praesent bibendum urna orci, a venenatis tellus venenatis at. Etiam ornare, est sed lacinia elementum, lectus diam tempor leo, sit amet elementum ex elit id ex. Ut ac viverra turpis. Quisque in nisl auctor, ornare dui ac, consequat tellus.

# Capítulo 4

## Research containing a figure

### 4.1. Introduction

This is the introduction. Sed vulputate tortor at nisl blandit interdum. Cras sagittis massa ex, quis eleifend purus condimentum congue. Maecenas tristique, justo vitae efficitur mollis, mi nulla varius elit, in consequat ligula nulla ut augue. Phasellus diam sapien, placerat sit amet tempor non, lobortis tempus ante.

### 4.2. Method

Donec imperdiet, lectus vestibulum sagittis tempus, turpis dolor euismod justo, vel tempus neque libero sit amet tortor. Nam cursus commodo tincidunt.

#### 4.2.1. SUBSECTION 1

This is the first part of the methodology. Duis tempor sapien sed tellus ultrices blandit. Sed porta mauris tortor, eu vulputate arcu dapibus ac. Curabitur sodales at felis efficitur sollicitudin. Quisque at neque sollicitudin, mollis arcu vitae, faucibus tellus.

#### 4.2.2. SUBSECTION 2

This is the second part of the methodology. Sed ut ipsum ultrices, interdum ipsum vel, lobortis diam. Curabitur sit amet massa quis tortor molestie dapibus a at libero. Mauris mollis magna quis ante vulputate consequat. Integer leo turpis, suscipit ac venenatis pellentesque, efficitur non sem. Pellentesque eget vulputate turpis. Etiam id nibh at elit fermentum interdum.

#### 4.3. Results

These are the results. In vitae odio at libero elementum fermentum vel iaculis enim. Nullam finibus sapien in congue condimentum. Curabitur et ligula et ipsum mollis fringilla.

#### 4.4. Discussion

Figure 4.1 shows how to add a figure. Donec ut lacinia nibh. Nam tincidunt augue et tristique cursus. Vestibulum sagittis odio nisl, a malesuada turpis blandit quis. Cras ultrices metus tempor laoreet sodales. Nam molestie ipsum ac imperdiet laoreet. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

#### 4.5. Conclusion

This is the conclusion to the chapter. Quisque nec purus a quam consectetur volutpat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. In lorem justo, convallis quis lacinia eget, laoreet eu metus. Fusce blandit tellus tellus. Curabitur nec cursus odio. Quisque tristique eros nulla, vitae finibus lorem aliquam quis. Interdum et malesuada fames ac ante ipsum primis in faucibus.

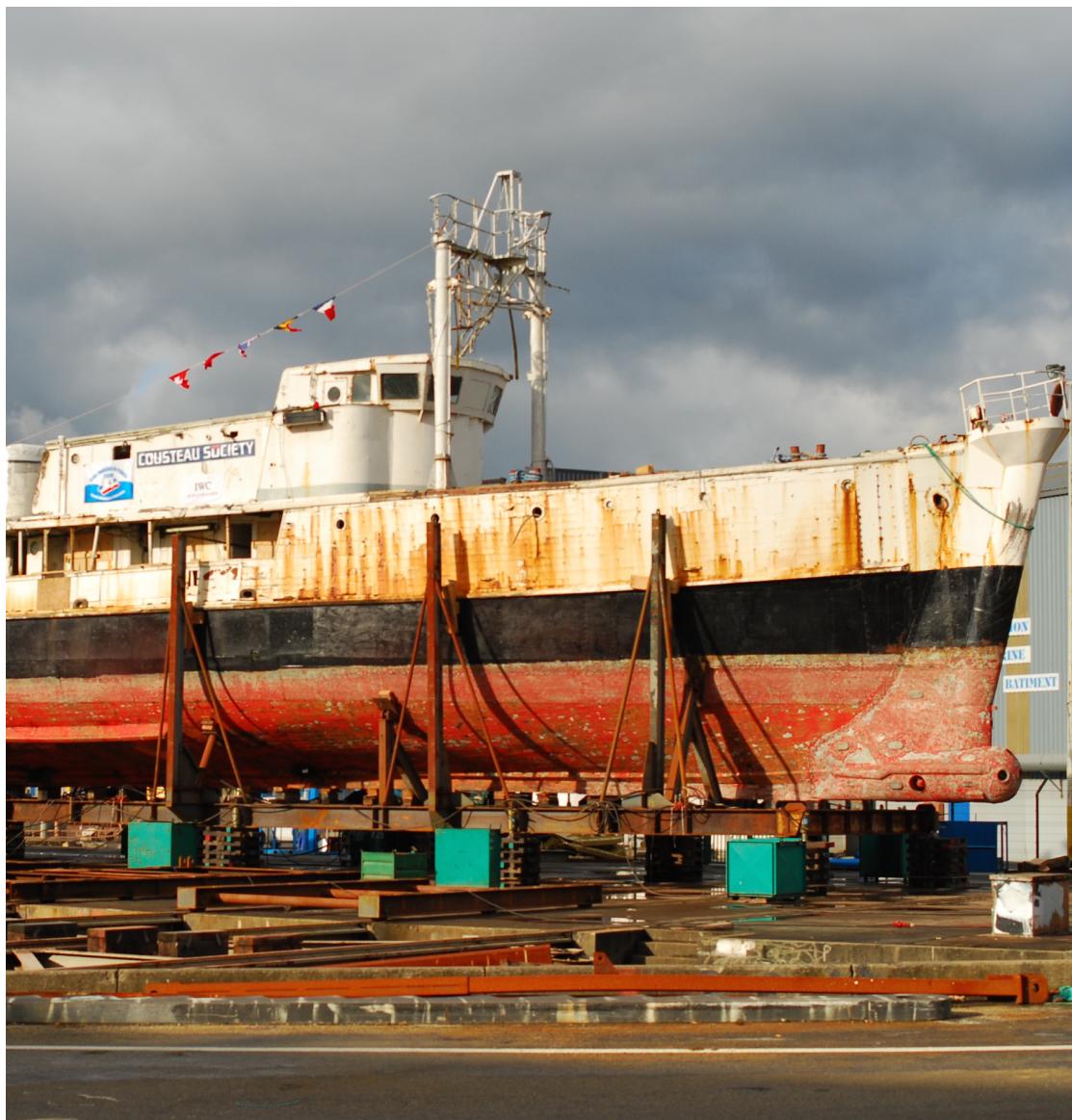


Figure 4.1: RV Calypso is a former British Royal Navy minesweeper converted into a research vessel for the oceanographic researcher Jacques-Yves Cousteau. It was equipped with a mobile laboratory for underwater field research.



Figure 4.2: This is not a boat

# Capítulo 5

## Research containing a table

### 5.1. Introduction

This is the introduction. Phasellus non purus id mauris aliquam rutrum vitae quis tellus. Maecenas rhoncus ligula nulla, fringilla placerat mi consectetur eu. Aenean nec metus ac est ornare posuere. Nunc ipsum lacus, gravida commodo turpis quis, rutrum eleifend erat. Pellentesque id lorem eget ante porta tincidunt nec nec tellus.

### 5.2. Method

Vivamus consectetur, velit in congue lobortis, massa massa lacinia urna, sollicitudin semper ipsum augue quis tortor. Donec quis nisl at arcu volutpat ultrices. Maecenas ex nibh, consequat ac blandit sit amet, molestie in odio. Morbi finibus libero et nisl dignissim, at ultricies ligula pulvinar.

### 5.2.1. SUBSECTION 1

This is the first part of the methodology. Integer leo erat, commodo in lacus vel, egestas varius elit. Nulla eget magna quam. Nullam sollicitudin dolor ut ipsum varius tincidunt. Duis dignissim massa in ipsum accumsan imperdiet. Maecenas suscipit sapien sed dui pharetra blandit. Morbi fermentum est vel quam pretium maximus.

### 5.2.2. SUBSECTION 2

This is the second part of the methodology. Nullam accumsan condimentum eros eu volutpat. Maecenas quis ligula tempor, interdum ante sit amet, aliquet sem. Fusce tellus massa, blandit id tempus at, cursus in tortor. Nunc nec volutpat ante. Phasellus dignissim ut lectus quis porta. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

## 5.3. Results

Table 5.1 shows us how to add a table. Integer tincidunt sed nisl eget pellentesque. Mauris eleifend, nisl non lobortis fringilla, sapien eros aliquet orci, vitae pretium massa neque eu turpis. Pellentesque tincidunt aliquet volutpat. Ut ornare dui id ex sodales laoreet.

Table 5.1: Important data for various land masses.

Landmass	% stuff	Number of Owls	Dolphins per Capita	How Many Foos	How Many Bars	Forbidden Float
North America	94 %	20,028	17,465	12,084	20,659	1.71
Central America	91 %	6564	6350	8,189	12,012	1.52
South America	86 %	3902	4127	5,205	6,565	1.28
Africa	84 %	2892	3175	3,862	4,248	1.1
Europe	92 %	20,964	17,465	15,303	24,203	1.58
Asia	87 %	6852	6350	8,255	11,688	1.47
Oceania	87 %	4044	4127	5,540	6,972	1.28
Antarctica	83 %	2964	3175	4,402	4,941	1.13

#### 5.4. Discussion

This is the discussion. As we saw in Table 5.1, many things are true, and other things are not. Etiam sit amet mi eros. Donec vel nisi sed purus gravida fermentum at quis odio. Vestibulum quis nisl sit amet justo maximus molestie. Maecenas vitae arcu erat. Nulla facilisi. Nam pretium mauris eu enim porttitor, a mattis velit dictum. Nulla sit amet ligula non mauris volutpat fermentum quis vitae sapien.

## 5.5. Conclusion

This is the conclusion to the chapter. Nullam porta tortor id vehicula interdum. Quisque pharetra, neque ut accumsan suscipit, orci orci commodo tortor, ac finibus est turpis eget justo. Cras sodales nibh nec mauris laoreet iaculis. Morbi volutpat orci felis, id condimentum nulla suscipit eu. Fusce in turpis quis ligula tempus scelerisque eget quis odio. Vestibulum et dolor id erat lobortis ullamcorper quis at sem.

# Capítulo 6

## Final research study

### 6.1. Introduction

This is the introduction. Nunc lorem odio, laoreet eu turpis at, condimentum sagittis diam. Phasellus metus ligula, auctor ac nunc vel, molestie mattis libero. Praesent id posuere ex, vel efficitur nibh. Quisque vestibulum accumsan lacus vitae mattis.

### 6.2. Method

In tincidunt viverra dolor, ac pharetra tellus faucibus eget. Pellentesque tempor a enim nec venenatis. Morbi blandit magna imperdiet posuere auctor. Maecenas in maximus est.

#### 6.2.1. SUBSECTION 1

This is the first part of the methodology. Praesent mollis sem diam, sit amet tristique lacus vulputate quis. Vivamus rhoncus est rhoncus tellus lacinia, a interdum sem egestas. Curabitur quis urna vel quam blandit semper vitae a leo.

Nam vel lectus lectus.

### 6.2.2. SUBSECTION 2

This is the second part of the methodology. Aenean vel pretium tortor. Aliquam erat volutpat. Quisque quis lobortis mi. Nulla turpis leo, ultrices nec nulla non, ullamcorper laoreet risus.

## 6.3. Results

These are the results. Curabitur vulputate nisl non ante tincidunt tempor. Aenean porta nisi quam, sed ornare urna congue sed. Curabitur in sapien justo. Quisque pulvinar ullamcorper metus, eu varius mauris pellentesque et. In hac habitasse platea dictumst. Pellentesque nec porttitor libero. Duis et magna a massa lacinia cursus.

## 6.4. Discussion

This is the discussion. Curabitur gravida nisl id gravida congue. Duis est nisi, sagittis eget accumsan ullamcorper, semper quis turpis. Mauris ultricies diam metus, sollicitudin ultricies turpis lobortis vitae. Ut egestas vehicula enim, porta molestie neque consectetur placerat. Integer iaculis sapien dolor, non porta nibh condimentum ut.

## 6.5. Conclusion

This is the conclusion to the chapter. Nulla sed condimentum lectus. Duis sed tempor erat, at cursus lacus. Nam vitae tempus arcu, id vestibulum sapien. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus.

# Capítulo 7

## Conclusion

### 7.1. Thesis summary

In summary, pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Nunc eleifend, ex a luctus porttitor, felis ex suscipit tellus, ut sollicitudin sapien purus in libero. Nulla blandit eget urna vel tempus. Praesent fringilla dui sapien, sit amet egestas leo sollicitudin at.

### 7.2. Future work

There are several potential directions for extending this thesis. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam gravida ipsum at tempor tincidunt. Aliquam ligula nisl, blandit et dui eu, eleifend tempus nibh. Nullam eleifend sapien eget ante hendrerit commodo. Pellentesque pharetra erat sit amet dapibus scelerisque.

Vestibulum suscipit tellus risus, faucibus vulputate orci lobortis eget. Nunc varius sem nisi. Nunc tempor magna sapien, euismod blandit elit pharetra sed. In dapibus magna convallis lectus sodales, a consequat sem euismod. Curabitur in

interdum purus. Integer ultrices laoreet aliquet. Nulla vel dapibus urna. Nunc efficitur erat ac nisi auctor sodales.

# **Appendix 1: Some extra stuff**

Add appendix 1 here. Vivamus hendrerit rhoncus interdum. Sed ullamcorper et augue at porta. Suspendisse facilisis imperdiet urna, eu pellentesque purus suscipit in. Integer dignissim mattis ex aliquam blandit. Curabitur lobortis quam varius turpis ultrices egestas.

## **Appendix 2: Some more extra stuff**

Add appendix 2 here. Aliquam rhoncus mauris ac neque imperdiet, in mattis eros aliquam. Etiam sed massa et risus posuere rutrum vel et mauris. Integer id mauris sed arcu venenatis finibus. Etiam nec hendrerit purus, sed cursus nunc. Pellentesque ac luctus magna. Aenean non posuere enim, nec hendrerit lacus. Etiam lacinia facilisis tempor. Aenean dictum nunc id felis rhoncus aliquam.

# References

- Golub, T.R. et al., 1999. Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. *Science*, 286(5439), pp.531–537. Available at: <https://www.science.org/doi/10.1126/science.286.5439.531> [Accessed September 18, 2023].
- Isabelle Guyon, S.G., 2004. Gisette. Available at: <https://archive.ics.uci.edu/dataset/170> [Accessed September 19, 2023].
- Ramaswamy, S. et al., 2001. Multiclass cancer diagnosis using tumor gene expression signatures. *Proceedings of the National Academy of Sciences*, 98(26), pp.15149–15154. Available at: <https://pnas.org/doi/full/10.1073/pnas.211566398> [Accessed September 18, 2023].