



UNIVERSITAT OBERTA DE CATALUNYA (UOC)  
MASTER'S DEGREE IN DATA SCIENCE (*Data Science*)

## MASTER'S THESIS

AREA: DATA MINING AND MACHINE LEARNING

# Study of the Transcriptional Function of Cyclin D1 in Leukemia xxx subtitle? xxx

---

Author: Antonio Milán Otero

Tutor: Carles Barceló

Professor: Jordi Casas

---

Barcelona, February 24, 2019



# Créditos/Copyright

Una página con la especificación de créditos/copyright para el proyecto (ya sea aplicación por un lado y documentación por el otro, o unificadamente), así como la del uso de marcas, productos o servicios de terceros (incluidos códigos fuente). Si una persona diferente al autor colaboró en el proyecto, tiene que quedar explicitada su identidad y qué hizo.

A continuación se ejemplifica el caso más habitual, aunque se puede modificar por cualquier otra alternativa:



Esta obra está sujeta a una licencia de Reconocimiento - NoComercial - SinObraDerivada  
3.0 España de Creative Commons.



# THESIS INDEX CARD

Title:	Study of the transcriptional function of Cyclin D1 in Leukemia
Author:	Antonio Milán Otero
Nombre del colaborador/a docente:	Carles Barceló
Nombre del PRA (PRA???):	Jordi Casas
Date of delivery (mm/aaaa):	06/2019
Degree or program:	MSc in Data Science
Thesis area:	Data Mining and Machine Learning
Language:	English
Keywords	To be completed ... 3 as maximum



# Dedication/Quote

Breves palabras de dedicatoria y/o una cita.





# Acknowledgment

Si se considera oportuno, mencionar a las personas, empresas o instituciones que hayan contribuido en la realización de este proyecto.



# Abstract

Texto con la síntesis del proyecto, esto es, un texto en el cual se explica de manera concisa la definición del proyecto/problema abordado, sus objetivos/métodos de resolución, y los resultados y conclusiones (no puede ser una lista, sino un texto continuo redactado de manera estructurada). Si es necesario poner una referencia en este texto, ésta será anotada a pie de la misma página. En este apartado se puede usar un lenguaje más literario y coloquial que para el resto del documento.

El Abstract se escribirá por duplicado. Una de las versiones tiene que ser **obligatoriamente en inglés**. La otra versión tiene que estar escrita en catalán o español. En caso de no escribir el resto del documento en inglés, será necesario escribir la segunda versión del Abstract en el idioma utilizado para el resto de la memoria. La palabra Abstract se cambiará por “**Resum**” o “**Resumen**” en la versión catalana y española, respectivamente.

Extensión recomendada: 250 palabras máximo.

Como escribir un buen Abstract (en inglés):

<http://www.ece.cmu.edu/koopman/essays/abstract.html>

**Palabras clave:** Keywords del trabajo separadas por comas. Por ejemplo para este documento podrían ser Modelo, Pauta, Plantilla, Memoria, Trabajo de Final de Grado/Máster



# Contents

Abstract	ix
Índice	xi
Llistado de Figuras	xiii
Listado de Tablas	1
<b>1 Introduction</b>	<b>3</b>
1.1 General description of the problem . . . . .	3
1.2 Motivation . . . . .	3
1.2.1 Why this project? . . . . .	3
1.2.2 What can I add? . . . . .	4
1.2.3 Personal interest . . . . .	4
1.2.4 How can this project improve my CV . . . . .	4
1.3 Project Objectives . . . . .	5
1.3.1 Primary . . . . .	5
1.3.2 Secondaries . . . . .	5
1.4 Description of the Methodology Used . . . . .	5
1.5 Project Research Plan . . . . .	5
<b>Bibliografía</b>	<b>5</b>



# List of Figures





## List of Tables



# Chapter 1

## Introduction

### 1.1 General description of the problem

La leucemia es un conjunto de procesos tumorales de que provoca un aumento descontrolado de leucocitos (glóbulos blancos) en la sangre u órganos linfáticos. La ciclina D1 es un oncogén frecuentemente sobreexpresado en cáncer sobretodo en Leucemia. Es conocido por ser uno de los principales reguladores el ciclo celular aunque se desconoce su función como regulador de la transcripción (proceso por el que se generan las proteínas que controlan todos los procesos celulares). Se sabe que la ciclina D1 se une a las regiones promotoras de muchos genes aunque se desconoce el resultado de su actividad transcripcional que se cree fundamental en el desarrollo de la Leucemia. Recientes estudios de análisis transcripcional global han generado una enorme cantidad de datos susceptibles de ser analizados por data mining y machine learning, siendo su interpretación fundamental para conocer los mecanismos básicos de la célula que pueden conducir a la Leucemia. Evidentemente, la generación de nuevos medicamentos dependerá de conocer bien estos procesos.

### 1.2 Motivation

#### 1.2.1 Why this project?

The amount of data collected in scientific researches has increased exponentially during the last decades, making the usage of data science methodologies a good fit for improving the final analysis and results. This project is a clear example of how the advances in data science can trigger new ways of doing science, expanding the existing tools in order to achieve better results.

### **1.2.2 What can I add?**

During the course of my master in data science, I have been learning about all the different aspects of a data science project, starting from a project management point of view and continuing with all the different phases of acquisition, storage, hypothesis and modeling, visualization and deployment. For this specific project, though, my focus will be in the area of data mining and machine learning, and that is what I think I can add to the project, my accumulated experience in the commented area.

### **1.2.3 Personal interest**

My personal interest in this project comes from the fact, or bad luck, of having close family and friends affected for Leukemia, therefore, as soon as I saw the proposal of this project I felt emotionally connected to it.

Apart from that first reason, I also consider that one of the best usages of the advances of data mining and machine learning is to help in the creation of a better society, being one of its strongest foundations the improvement of the quality of the health of each individual. Therefore, I feel responsible of using my new acquired knowledge in areas that can lead to that goal.

### **1.2.4 How can this project improve my CV**

At the time of writing this, I am working as a research engineer in the control software group in MAX IV Laboratory, a synchrotron located in the south of Sweden that has the purpose of improving the scientific researches in a global encompass. Until now, my main activities has been related to the build of software for all the different aspects of the control system, from the very low level control, writing drivers for equipments, up to the high level software like graphical user interfaces that enable the scientist to perform their jobs, passing through all the layers in between, like software libraries, servers, etc. In other words, I have been always close to the control, synchronization and data acquisition, with this project, I can expand my coverage and help also in the next phase of a scientific research, the analysis of the generated data.

## **1.3 Project Objectives**

### **1.3.1 Primary**

### **1.3.2 Secondaries**

## **1.4 Description of the Methodology Used**

## **1.5 Project Research Plan**



# Bibliography