



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería Informática



TFG del Grado en Ingeniería
Informática

Simulador árboles de decisión



Presentado por Daniel Drefs Fernandes
en Universidad de Burgos — 22 de marzo
de 2024

Tutor: Carlos López Nozal, Ismael Ramos
Pérez



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería Informática



D. nombre tutor, profesor del departamento de nombre departamento, área de nombre área.

Expone:

Que el alumno D. Daniel Drefs Fernandes, con DNI dni, ha realizado el Trabajo final de Grado en Ingeniería Informática titulado título de TFG.

Y que dicho trabajo ha sido realizado por el alumno bajo la dirección del que suscribe, en virtud de lo cual se autoriza su presentación y defensa.

En Burgos, 22 de marzo de 2024

Vº. Bº. del Tutor:

Vº. Bº. del co-tutor:

D. nombre tutor

D. nombre co-tutor

Resumen

En este primer apartado se hace una **breve** presentación del tema que se aborda en el proyecto.

Descriptores

Palabras separadas por comas que identifiquen el contenido del proyecto Ej: servidor web, buscador de vuelos, android ...

Abstract

A **brief** presentation of the topic addressed in the project.

Keywords

keywords separated by commas.

Índice general

Índice general	iii
Índice de figuras	iv
Índice de tablas	v
1. Introducción	1
2. Objetivos del proyecto	3
3. Conceptos teóricos	5
3.1. Decision Trees	5
3.2. Referencias	6
3.3. Imágenes	6
3.4. Listas de items	7
3.5. Tablas	7
4. Técnicas y herramientas	9
5. Aspectos relevantes del desarrollo del proyecto	11
6. Trabajos relacionados	13
7. Conclusiones y Líneas de trabajo futuras	15
Bibliografía	17

Índice de figuras

3.1. Decision tree example	6
3.2. Autómata para una expresión vacía	7

Índice de tablas

3.1. Herramientas y tecnologías utilizadas en cada parte del proyecto	8
---	---

1. Introducción

Decision trees are powerful tools in machine learning and data mining, serving as intuitive and interpretable models for decision-making. Two notable algorithms, IDE3 and C4.5, have significantly contributed to their development and widespread use. IDE3, or Iterative Dichotomiser 3, laid the foundation for decision tree learning by recursively partitioning data based on attribute values, aiming to maximize information gain at each step. C4.5 improved upon IDE3 by handling continuous attributes, missing values, and pruning techniques, enhancing the robustness and accuracy of decision trees. These algorithms play a crucial role in various fields, enabling efficient classification and prediction tasks while offering insights into decision-making processes.

That is why this project aims to build a simulator in form of a web application that teaches these algorithms in a way that is easy to understand.

2. Objetivos del proyecto

The primary objective of this project has been to create an interactive and informative web application focused on educating users about decision tree algorithms such as IDE3 and C4.5. In relation to that, it also teaches the concept of entropy and how it is connected to decision trees. The users are provided with explanations, color-coded and dynamic step-by-step visualizations which make the concepts easier to understand. The server uses SVG images to display the algorithm's progress and entropy functions.

3. Conceptos teóricos

In the following, all the theoretical concepts relevant for the understanding of the project will be explained.

Algunos conceptos teóricos de \LaTeX ¹.

3.1. Decision Trees

A decision tree [2] is a versatile supervised learning algorithm used for classification and regression tasks. Its goal is to predict the value of a variable based on previously processed input. Its hierarchical structure includes a root node and several internal and leaf nodes.

It starts at the root, which represents the feature that best separates the underlying dataset based on a certain criterion. An example would be information gain, which will be explained in a later section. From there branches extend to internal nodes, also called decision nodes. These internal nodes also represent features along with a decision rule that tells us how to further split the data. These features are continuously evaluated until homogenous subsets are created by the leaf nodes. These represent all the possible outcomes of the dataset with each one corresponding to a class label.

Decision tree learning utilizes a divide and conquer approach, iteratively finding the best split points until all or most of the input data is classified.

An example of a decision tree which evaluates whether a person is going to be an astronaut or not:

¹Créditos a los proyectos de Álvaro López Cantero: Configurador de Presupuestos y Roberto Izquierdo Amo: PLQuiz

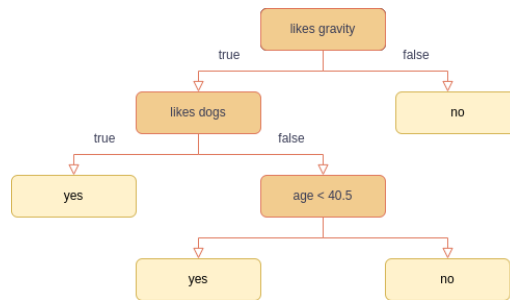


Figura 3.1: Decision tree example

Subsecciones

Además de secciones tenemos subsecciones.

Subsubsecciones

Y subsecciones.

3.2. Referencias

Las referencias se incluyen en el texto usando cite [?]. Para citar webs, artículos o libros [3], si se desean citar más de uno en el mismo lugar [1, 3].

3.3. Imágenes

Se pueden incluir imágenes con los comandos standard de \LaTeX , pero esta plantilla dispone de comandos propios como por ejemplo el siguiente:



Figura 3.2: Autómata para una expresión vacía

3.4. Listas de items

Existen tres posibilidades:

- primer item.
- segundo item.

1. primer item.
2. segundo item.

Primer item más información sobre el primer item.

Segundo item más información sobre el segundo item.

▪

3.5. Tablas

Igualmente se pueden usar los comandos específicos de \LaTeX o bien usar alguno de los comandos de la plantilla.

Herramientas	App	AngularJS	API REST	BD	Memoria
HTML5		X			
CSS3		X			
BOOTSTRAP		X			
JavaScript		X			
AngularJS		X			
Bower		X			
PHP			X		
Karma + Jasmine		X			
Slim framework			X		
Idiorm			X		
Composer			X		
JSON		X	X		
PhpStorm		X	X		
MySQL				X	
PhpMyAdmin				X	
Git + BitBucket		X	X	X	X
MikTeX					X
TeXMaker					X
Astah					X
Balsamiq Mockups		X			
VersionOne		X	X	X	X

Tabla 3.1: Herramientas y tecnologías utilizadas en cada parte del proyecto

4. Técnicas y herramientas

Esta parte de la memoria tiene como objetivo presentar las técnicas metodológicas y las herramientas de desarrollo que se han utilizado para llevar a cabo el proyecto. Si se han estudiado diferentes alternativas de metodologías, herramientas, bibliotecas se puede hacer un resumen de los aspectos más destacados de cada alternativa, incluyendo comparativas entre las distintas opciones y una justificación de las elecciones realizadas. No se pretende que este apartado se convierta en un capítulo de un libro dedicado a cada una de las alternativas, sino comentar los aspectos más destacados de cada opción, con un repaso somero a los fundamentos esenciales y referencias bibliográficas para que el lector pueda ampliar su conocimiento sobre el tema.

5. Aspectos relevantes del desarrollo del proyecto

Este apartado pretende recoger los aspectos más interesantes del desarrollo del proyecto, comentados por los autores del mismo. Debe incluir desde la exposición del ciclo de vida utilizado, hasta los detalles de mayor relevancia de las fases de análisis, diseño e implementación. Se busca que no sea una mera operación de copiar y pegar diagramas y extractos del código fuente, sino que realmente se justifiquen los caminos de solución que se han tomado, especialmente aquellos que no sean triviales. Puede ser el lugar más adecuado para documentar los aspectos más interesantes del diseño y de la implementación, con un mayor hincapié en aspectos tales como el tipo de arquitectura elegido, los índices de las tablas de la base de datos, normalización y desnormalización, distribución en ficheros³, reglas de negocio dentro de las bases de datos (EDVHV GH GDWRV DFWLYDV), aspectos de desarrollo relacionados con el WWW... Este apartado, debe convertirse en el resumen de la experiencia práctica del proyecto, y por sí mismo justifica que la memoria se convierta en un documento útil, fuente de referencia para los autores, los tutores y futuros alumnos.

6. Trabajos relacionados

Este apartado sería parecido a un estado del arte de una tesis o tesina. En un trabajo final grado no parece obligada su presencia, aunque se puede dejar a juicio del tutor el incluir un pequeño resumen comentado de los trabajos y proyectos ya realizados en el campo del proyecto en curso.

7. Conclusiones y Líneas de trabajo futuras

Todo proyecto debe incluir las conclusiones que se derivan de su desarrollo. Éstas pueden ser de diferente índole, dependiendo de la tipología del proyecto, pero normalmente van a estar presentes un conjunto de conclusiones relacionadas con los resultados del proyecto y un conjunto de conclusiones técnicas. Además, resulta muy útil realizar un informe crítico indicando cómo se puede mejorar el proyecto, o cómo se puede continuar trabajando en la línea del proyecto realizado.

Bibliografía

- [1] Zachary J Bortolot and Randolph H Wynne. Estimating forest biomass using small footprint lidar data: An individual tree-based approach that incorporates training data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59(6):342–360, 2005.
- [2] IBM. What is a decision tree? <https://www.ibm.com/topics/decision-trees#:~:text=A%20decision%20tree%20is%20a,internal%20nodes%20and%20leaf%20nodes.>, -. [Internet; visitado 15-marzo-2024].
- [3] John R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, 1992.