



Eidgenössische Technische Hochschule Zürich  
Swiss Federal Institute of Technology Zurich

*Repartido*

*Informática*



# Recetas Inteligentes

Tesis de licenciatura

Alessio Bähler

[abaehler@student.ethz.ch](mailto:abaehler@student.ethz.ch)

Grupo de Computación Distribuida  
Laboratorio de Ingeniería Informática y Redes  
ETH Zurich

Supervisores:

laura par

Prof. Dr. Roger Wattenhofer

22 de diciembre de 2015

# Resumen

Con la creciente popularidad de los sitios web para compartir recetas en los últimos años, Internet se ha convertido en una fuente accesible de grandes cantidades de información sobre recetas. Un problema importante que encontramos es que el formato de la receta varía mucho entre las diferentes fuentes. Por lo tanto, **recopilamos una gran cantidad de recetas de diferentes sitios web y analizamos el contenido para obtener una estructura uniforme**. Posteriormente, **implementamos un esquema de clasificación para determinar el nivel de dificultad de una receta en función de los datos recopilados**. Para este propósito, consideramos los ingredientes, los tiempos y las instrucciones, y utilizamos las técnicas de clasificación **Gaussian Naive Bayes y Decision Tree** para evaluar la calidad de nuestros métodos. Además, consideramos el espacio determinado por los ingredientes más frecuentes que se encuentran entre un subconjunto de recetas para descubrir similitudes y diferencias entre recetas pertenecientes a diferentes cocinas del mundo y categorías de comidas, por ejemplo, platos principales, postres o aperitivos.

# Contenido

Resumen	i
1. Introducción	1
1.1 Motivación . . . . .	1
1.2 Trabajo relacionado . . . . .	2
1.3 Descripción general . . . . .	2
2. Fondo	3
2.1 Clasificación . . . . .	3
2.1.1 Bayesiano ingenuo gaussiano . . . . .	3
2.1.2 Árboles de clasificación . . . . .	4
2.2 Agrupación de K-Means . . . . .	5
3 Diseño	7
3.1 Adquisición de datos . . . . .	7
3.1.1 Recopilación de recetas . . . . .	7
3.1.2 Estructuración de los datos recopilados . . . . .	8
3.2 Clasificación de dificultad de la receta . . . . .	10
3.2.1 Desarrollo de un método objetivo . . . . .	10
3.2.2 Evaluando la validez de nuestros métodos . . . . .	11
3.3 Descubriendo similitudes de recetas . . . . .	11
3.3.1 Similitud de recetas utilizando la agrupación en clústeres de K-Means . . . . .	11
3.3.2 Similitud de recetas utilizando árboles de clasificación . . . . .	12
4 Implementación	13
5 resultados	14
5.1 Resultados de la clasificación de dificultad . . . . .	14
5.2 Similitudes de categoría . . . . .	dieciséis

Contenido	iii
5.2.1 Agrupación de K-Means . . . . .	dieciséis
5.2.2 Clasificación de árboles . . . . .	21
5.3 Similitudes culinarias . . . . .	26
5.3.1 Agrupación de K-Means . . . . .	26
5.3.2 Clasificación de árboles . . . . .	30
6 Perspectiva y Resumen	34
Bibliografía	35
7 Apéndice Capítulo	37

# Introducción

---

## 1.1 Motivación

Desde la antigüedad, registrar y compartir recetas de cocina ha sido una actividad humana tradicional. La descripción de la receta generalmente se imprimía en una hoja de papel. La creciente ubicuidad y popularidad de Internet abrió las puertas para compartir recetas en colaboración a gran escala como nunca antes había sido posible.

Actualmente, existen millones de recetas de todo el mundo, publicadas gratuitamente por personas de diferentes culturas y con diferentes habilidades culinarias que son accesibles para cualquier persona. Esta tendencia es ventajosa ya que permite el intercambio de recetas entre lugares geográficos y culturas mutuamente distantes, pero también tiene un gran inconveniente: es difícil encontrar recetas específicas entre millones de personas.

Por lo tanto, estamos interesados en encontrar formas de mejorar la búsqueda de recetas por diferentes parámetros como ingredientes, cocina, categoría, valores nutricionales, costo, etc.

Pero podríamos hacer mucho más. **Podríamos encontrar formas de combinar ingredientes para crear nuevas recetas, reemplazar ingredientes con otros similares o aprender más sobre los hábitos alimenticios de las personas y brindar consejos útiles sobre cómo mejorar la calidad de los alimentos.** Dado que hoy en día incluso es posible comprar alimentos en Internet, podríamos construir un servicio que, por ejemplo, genere recetas para el usuario y le permita comprar todos los ingredientes necesarios con un solo clic. Posteriormente, los productos podrían ser enviados directamente al domicilio del usuario.

El primer paso que tenemos que dar es, obviamente, obtener los datos. Sin embargo, no existe una estructura común ni un lenguaje común entre los diferentes repositorios de recetas, por lo tanto, los datos recopilados deben procesarse y estructurarse de manera uniforme antes de poder realizar cualquier análisis. En este trabajo, damos los primeros pasos para recopilar una gran cantidad de recetas de diferentes sitios web para compartir recetas y guardamos cada una en una estructura uniforme. Además, tratamos de desarrollar un método para determinar el nivel de dificultad de una receta observando solo los ingredientes que se necesitan, los tiempos y las instrucciones. Para evaluar la precisión utilizamos algunos algoritmos de clasificación simples. También utilizamos la agrupación de K-Means para encontrar relaciones entre recetas pertenecientes a diferentes cocinas regionales o categorías de comidas en el espacio definido por los ingredientes más frecuentes.

## 1.2 Trabajo relacionado

Las características que ofrecen los sitios web para compartir recetas varían desde una simple lista de recetas a opciones de búsqueda avanzada y, en ocasiones, incluso a la gestión de listas de la compra. El sitio web [allrecipes.com](#) [1], por ejemplo, agrupa recetas en cocina y categoría, por ejemplo plato principal y aperitivo. Adicionalmente, permite la búsqueda de nombres de recetas e ingredientes, ofrece listas de verificación de ingredientes e incluso permite al usuario comprar algunos productos en línea. En contraste, [epicurious.com](#) [2] no ofrece las dos últimas opciones, pero tiene capacidades de búsqueda ampliadas, que permiten al usuario filtrar los resultados por comida, dieta, ingredientes, cocina, tipo de plato y método de preparación.

La recuperación y recomendación de recetas también ha sido un tema de investigación común.

El trabajo anterior incluye considerar ingredientes superpuestos para encontrar recetas similares. basado en el historial de navegación web relacionado con la cocina [3] o en las calificaciones de recetas anteriores de los usuarios [4]. Teng et al. [5] desarrolló dos redes de ingredientes diferentes que expresan relaciones entre ingredientes que ocurren juntos con frecuencia, así como candidatos de sustitución para predecir las calificaciones de recetas por parte de los usuarios. Otra rama de investigación se centró en descubrir relaciones entre ingredientes y cocinas utilizando modelos probabilísticos generativos [6] y técnicas de clasificación [7].

## 1.3 Resumen

En el Capítulo 2 presentamos los diferentes algoritmos de aprendizaje automático utilizados a lo largo de este trabajo. En el Capítulo 3 mostramos el proceso de recolección y estructuración (Sección 3.1), cómo desarrollamos nuestro modelo de clasificación de dificultad (Sección 3.2) y, finalmente, cómo funciona el agrupamiento espacial de ingredientes (Sección 3.3). Capítulo 4 contiene una breve descripción general de las herramientas que usamos para implementar los scripts y las bases de datos, mientras que los resultados del análisis se explican en el Capítulo 5. Capítulo 6 sirve como resumen de nuestro trabajo y propone algunas posibles ampliaciones.

# Fondo

---

Este capítulo pretende dar una introducción rápida a los diferentes algoritmos de aprendizaje automático utilizados durante este trabajo. En primer lugar, explicamos los enfoques de clasificación de Gaussian Naive Bayes y Decision Trees y, en segundo lugar, el enfoque de agrupamiento de K-Means.

## 2.1 Clasificación

Los algoritmos de aprendizaje automático se dividen en diferentes categorías, una de las cuales es el aprendizaje supervisado. En el aprendizaje supervisado, la tarea es inferir una función a partir de un conjunto de datos etiquetado como "entrenamiento". El conjunto de datos contiene ejemplos en forma de tuplas, donde el primer elemento representa la entrada y el segundo representa la salida deseada, que se alimentan al algoritmo. El algoritmo usa estos datos para inferir una función que luego se puede usar para predecir la salida de valores no incluidos en el conjunto de datos de entrenamiento, es decir, cuya salida no se conoce. La clasificación es el subconjunto de técnicas de aprendizaje supervisado que considera distintas categorías como resultado objetivo.

### 2.1.1 Bayesiano ingenuo gaussiano

La clasificación Naive Bayes se refiere a un conjunto de algoritmos de clasificación basados en la aplicación del teorema de Bayes con la suposición "ingenua" de que todas las variables son independientes por pares. Usando diferentes distribuciones es posible obtener diferentes algoritmos.

Dada una variable de categoría  $y$  y un número de características  $x_1, \dots, x_n$ , el teorema de Bayes establece que la probabilidad condicional de que  $y$  sea un cierto valor  $Y$  y el valor de cada  $x_1, \dots, x_n$ , se puede descomponer como  $P(y)P(x_1, \dots, x_n|y)$

$$P(y|x_1, \dots, x_n) = \frac{P(y)P(x_1, \dots, x_n|y)}{P(x_1, \dots, x_n)},$$

Con la suposición ingenua de que todas las características  $x_1, \dots, x_n$  son independientes por pares, podemos aplicar repetidamente la regla de independencia  $P(x_i, x_j|y) = P(x_i|y) \cdot P(x_j|y)$

## 2. Fondo

4

en  $P(x_1, \dots, x_n)$  y obtener

$$P(y|x_1, \dots, x_n) = \frac{P(y) \prod_{i=1}^n P(x_i|y)}{P(x_1, \dots, x_n)}.$$

Dado que  $P(x_1, \dots, x_n)$  no depende de  $y$ , sino solo de los valores de entrada  $x_1, \dots, x_n$ , para encontrar la clase que se ajuste con mayor probabilidad a los datos de entrada, podemos usar

$$y = \underset{y}{\operatorname{argmax}} P(y) \prod_{i=1}^n P(x_i|y)$$

como regla de clasificación.

En Gaussian Naive Bayes, la función de distribución de probabilidad  $P$  es

$$P(x_i|y) = \frac{1}{\sigma_y \sqrt{2\pi}} \exp - \frac{(x_i - \mu_i)^2}{2\sigma_y^2}.$$

## 2.1.2 Árboles de clasificación

Los métodos de aprendizaje automático basados en árboles dividen recursivamente el espacio de características en un conjunto de espacios y luego ajustan un modelo simple en cada uno. La partición se representa mediante una estructura de árbol compuesta por nodos de decisión internos y hojas terminales. Cada nodo de decisión implementa una función de prueba con resultados discretos que etiquetan las ramas. Para predecir el resultado de una entrada, se recorre el árbol desde la raíz hasta llegar a una hoja, en cuyo punto la categoría asignada a la hoja constituye la salida. Durante este proceso se aplica una prueba en cada nodo y dependiendo del resultado se toma una de las ramas. La figura 2.1 muestra un ejemplo de cómo funcionan los algoritmos en dos dimensiones. Las características  $x_1$  y  $x_2$  deben utilizarse para distinguir los objetos de la categoría  $C_1$  de los objetos de la categoría  $C_2$ . En primer lugar, el espacio se divide en dos regiones mediante una línea vertical correspondiente al valor  $w_{10}$  de la característica  $x_1$  y, en segundo lugar, la región derecha se vuelve a dividir en dos, pero utilizando una línea horizontal correspondiente al valor  $w_{20}$  de la característica  $x_2$ . Cada una de las regiones resultantes contiene solo un tipo o clase de objeto y, por lo tanto, podemos construir un árbol de decisiones que asigne objetos a una categoría comparando sus coordenadas  $(x_1, x_2)$  con  $w_{10}$  y  $w_{20}$ . Por ejemplo, un objeto con una coordenada  $x_1$  menor que  $w_{10}$  caerá a la izquierda de la línea vertical, por lo tanto, se asignará a  $C_1$  porque la condición en el nodo raíz no se cumple.

Para fines de clasificación, se utiliza una medida de impurezas para cuantificar la bondad de una región dividida, determinando así la forma del árbol. Se dice que una división es pura si después de la división, para todas las ramas, todas las instancias que eligen una rama pertenecen a la misma clase. En tal caso, no hay necesidad de dividir más y se puede agregar al árbol un nodo de hoja con la etiqueta correspondiente. Si un nodo no es puro, las instancias deben derramarse para disminuir la impureza hasta que se alcance un criterio de parada.



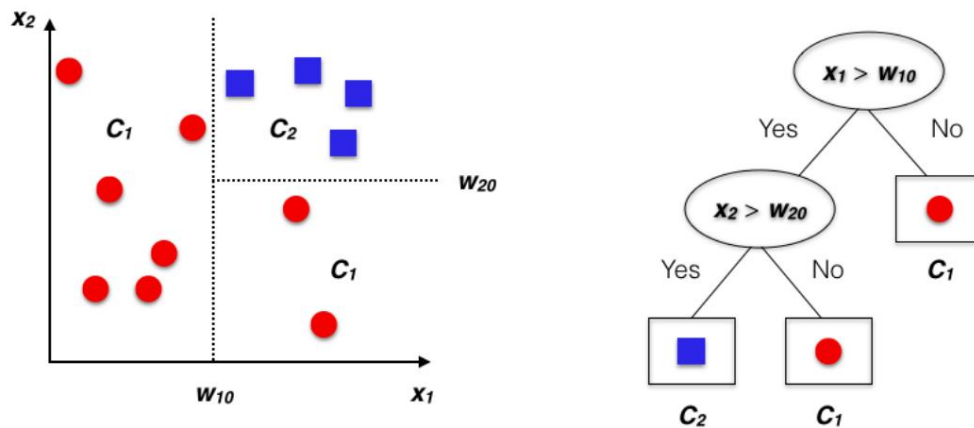


Figura 2.1: Ejemplo de construcción del Árbol de Clasificación. A la izquierda vemos el espacio separado en tres regiones diferentes. A la derecha vemos el árbol binario correspondiente.

## 2.2 Agrupación de K-Means

El objetivo de los algoritmos de agrupamiento es dividir un conjunto de puntos de datos observados en grupos, que se denominan conglomerados, de modo que las diferencias por pares entre los asignados al mismo conglomerado tienden a ser menores que las de los puntos de datos ubicados en diferentes conglomerados [8]. El agrupamiento de K-Means es uno de esos algoritmos. El número de conglomerados debe conocerse a priori y la similitud entre los puntos de datos se expresa utilizando alguna forma de métrica de distancia. Un ejemplo de ello es la distancia euclidiana al cuadrado. Denotamos la distancia entre los puntos  $a$  y  $b$  con  $d(a, b)$ . Cada grupo  $C$  está entonces representado por la media  $\mu_C$ , que está definida por

$$\mu_C = \frac{1}{|C|} \sum_{x \in C} x, \quad (2.1)$$

donde  $|C|$  es el número de puntos dentro del grupo  $C$ . Debido a que la media representa el punto "medio" del grupo, también se le llama centroide. El objetivo es encontrar los centroides que minimicen la suma de todas las distancias al cuadrado de todas las observaciones dentro de cada grupo. Es decir, para cada grupo queremos encontrar una media  $\mu$  tal que

$$\mu = \underset{\mu}{\text{mínimo de argumento}} \sum_{x \in C} d(x, \mu). \quad (2.2)$$

Como se ve en el Algoritmo 1, el primer paso consiste en seleccionar aleatoriamente un centroide para cada grupo. El siguiente paso asigna cada punto de datos al centroide más cercano. Luego, todos los centroides se actualizan calculando la media de todos los puntos de datos observados que se les asignan mediante la fórmula 2.1. La diferencia entre el

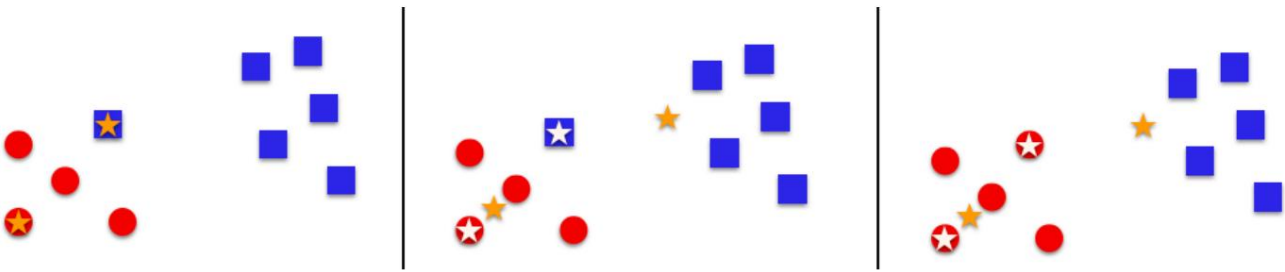


Figura 2.2: Ejemplo de cómo funciona el algoritmo de agrupación en clústeres de K-Means utilizando dos clústeres en un conjunto de puntos de datos bidimensionales. En primer lugar, se eligen aleatoriamente dos puntos como centroides iniciales (estrellas amarillas) y los puntos se asignan al centroide más cercano. En segundo lugar, se calcula la media dentro del conglomerado y se utiliza para actualizar los centroides. Por último, los puntos se reasignan al nuevo centroide más cercano.

Luego, los valores nuevos y antiguos de los centroides se calculan y comparan con un umbral para determinar si el algoritmo debe continuar o puede detenerse porque los centroides no se movieron significativamente. Si no se alcanza este umbral, nuevamente asignamos puntos de datos a los centroides de clúster recién establecidos.

---

#### Algoritmo 1: Agrupación de K-medias

---

1. Los centroides se inicializan con puntos aleatorios del conjunto de datos.
  2. Cada otro punto en el conjunto de datos se asigna a su centroide más cercano de acuerdo con la medida de distancia elegida.
  3. Para cada grupo, se calcula la media de todos los puntos asignados y se convierte en el nuevo centroide.
  4. Si el conteo de iteraciones excedió el número máximo de bucles o se cumple el criterio de convergencia dado, finalice el algoritmo; de lo contrario, ejecute el paso 2.
-

## Capítulo 3

## Diseño

En este capítulo nos fijamos en las dos fases que se pueden distinguir en este trabajo: en primer lugar, describimos métodos para recopilar datos de recetas estructurados y, en segundo lugar, discutir nuestros esquemas de análisis.

### 3.1 Adquisición de datos

#### 3.1.1 Recopilación de recetas

Un gran número de sitios web ofrecen recetas de cocina de cualquier tipo. A veces ellos son publicados por cocineros profesionales, a veces por usuarios con diversos grados de pericia. Rastreamos los sitios web más populares en tres idiomas diferentes: Inglés, alemán e italiano. Se tomó esta elección porque las diferentes culturas expresar diferentes estilos de cocina y, aunque gracias a Internet las barreras espaciales hoy en día son fáciles de romper, la barrera del idioma sigue ahí. También el intercambio de recetas entre cocinas no está exento de modificaciones, por ejemplo un plato de lasaña probablemente se cocinará de manera diferente en los EE. UU. en comparación con una lasaña italiana tradicional.


En total, se recopilaron alrededor de 1 millón de recetas de los siguientes sitios web:

Inglés	alemán	italiano
www.allrecipes.com	www.chefkoch.de	www.giallozafferano.it
www.bbc.com	www.lecker.de	www.repubblica.it
www.chow.com		www.williamssonoma.com
www.epicurious.com		www.ricettedellanonna.it
www.food.com		www.saleepepe.it
www.foodnetwork.com		
www.cucharada.com		
608136	327806	14750

### 3.1.2 Estructuración de los datos recopilados

Dado que la mayoría de los sitios web muestran las recetas como texto no estructurado, es necesario analizar las páginas web para extraer cualquier información posible que pueda ser necesaria posteriormente para nuestro análisis. Ejemplos sencillos son los ingredientes, las direcciones y los tiempos. Como no sabíamos de antemano toda la información que necesitaríamos de cada receta, primero descargamos el código HTML completo de cada página web y solo después analizamos el código y extrajimos la información buscada. En este proceso, se tuvo en cuenta la estructura de cada sitio web y se atravesó el árbol DOM de cada página web en consecuencia.

La Figura 3.1 muestra un ejemplo de cómo se ve una página web, mientras que la Figura 3.2 muestra la estructura HTML del sitio web.



## Bolognese Lasagna

Prep Time: 30 minutes   Cook Time: 60 minutes   Servings: 12

Make a big batch of Bolognese sauce and use it to prepare this hearty lasagna.

[Add to Recipe Box](#)   [Email](#)   [Print](#)

[G+1](#)   [67](#)   [Tweet](#)   [Pin it](#)

**Ingredients:**

- 16 lasagna sheets
- 4 cups ricotta cheese
- 2 eggs
- 1/4 cup finely chopped fresh flat-leaf parsley
- Salt and freshly ground pepper, to taste
- 6 cups [Bolognese sauce](#)
- 1 1/2 lb. mozzarella cheese, thinly sliced
- 2 cups grated Parmigiano-Reggiano cheese
- 3 Tbs. slivered fresh basil

**Related Recipes**

[Bolognese Sauce](#) ▶

**Directions:**

Preheat an oven to 375°F. Grease a 9-by-13-inch lasagna pan or baking dish with olive oil.

Line a baking sheet with paper towels. Bring a large pot of salted water to a boil over high heat. Add the lasagna sheets and cook according to the package instructions. Drain, rinse with cold water and drain again. Transfer to the prepared baking sheet and pat dry.

In a bowl, stir together the ricotta, eggs, parsley, salt and pepper.

Spread 1/4 cup of the Bolognese sauce in the bottom of the prepared pan. Lay 4 lasagna sheets, slightly overlapping, on the sauce. Spread 1 cup of the ricotta mixture evenly over the pasta. Spread one-fourth of the remaining sauce over the ricotta mixture. Arrange one-fourth of the mozzarella slices on the sauce, then sprinkle with 1/2 cup of the Parmigiano-Reggiano. Lay 4 more lasagna sheets, slightly overlapping, on top, and layer with another portion of the ricotta mixture, sauce, mozzarella and Parmigiano-Reggiano. Repeat the layering 2 more times, for a total of 4 layers. Sprinkle the basil evenly on top.

Cover the pan with aluminum foil and bake for 25 minutes. Uncover the pan and continue baking until the sauce is bubbling, 25 to 30 minutes more. Remove the pan from the oven and let rest, uncovered, for 15 minutes before serving. Serves 10 to 12.

Figura 3.1: Ejemplo de página web de recetas de williams-sonoma.com

La figura 3.4 muestra algunos ejemplos de los ingredientes más frecuentes encontrados. Una descrip-

Campo	Contenido
Descripción	Descripción de la receta
Nombre	nombre de la receta
Autor	Nombre del usuario que publicó la receta
Ingredientes	Lista de ingredientes con su cantidad
tiempo de preparación	Tiempo de preparación
hora de cocinar	Hora de cocinar
tiempo	Tiempo Total
Direcciones	Lista de pasos con instrucciones
Recetas relacionadas	Urls a otras recetas mostradas en la misma página web
notas	Sugerencias o comentarios del autor.
Reseñas	Comentarios de otros usuarios

Figura 3.2: Nombre del campo y descripción del contenido de una receta estructurada.

ción del mismo ingrediente podría de hecho diferir de muchas maneras diferentes, pero la mayoría comúnmente encontramos la presencia de adjetivos o nombres de marcas y abreviaturas para ingredientes bien conocidos. Para abordar el problema, realizamos un análisis de N-gramas en todos los nombres de los ingredientes. Un N-grama es una secuencia de N palabras dentro de un frase. La Figura 3.3 muestra todos los gramos de la "Crema de champiñones condensada". Calculamos las frecuencias de todos los 1, 2, 3 y 4 gramos que se encuentran en todas las recetas en ingredientes al dividir todas las cadenas de ingredientes en palabras separadas. Así recopilamos una gran cantidad de estos N-gramas que almacenamos en nuestra base de datos junto con sus Frecuencia de ocurrencia. Obviamente no todos los gramos computados son ingredientes nombres, por lo tanto eliminamos adjetivos y marcas de los ingredientes antes de calcular los gramos para reducir el espacio de búsqueda. Luego filtramos la base de datos eliminando nombres de ingredientes no significativos y fusionando diferentes gramos que se referían al mismo ingrediente, como nombres en plural y singular.

"Crema condensada de champiñones"	
1 gramo	"Condensado", "crema", "de", "champiñones", "sopa"
2 gramos de	"crema condensada", "crema de", "de champiñones", "sopa de champiñones"
3 gramos de	"crema condensada de", "crema de champiñones", "de sopa de champiñones"
4 gramos de	"crema de champiñones condensada", "crema de champiñones"
5 gramos de	"Crema de champiñones condensada"
6 gramos {}	

Figura 3.3: N-gramas de la cadena "Crema de champiñones condensada"

Recuento	Ingrediente	Diferentes descripciones
20142	sal	sal marina, sal marina hawaiana, sal marina molida
18674	pimienta	pimienta, pimienta limón, pimienta blanca, pimienta sazónada
18015	mantequilla	azúcar, azúcar glas, azúcar granulado, azúcar superfino
13204	de azúcar	mantequilla, mantequilla, blanda, mantequilla sin sal, congelada
12485	aceite	aceite, aceite para freír, aceite vegetal, aceite de oliva, aceite de sésamo oscuro
11345	queso	mozzarella, queso mozzarella, queso cheddar en cubos
10864	harina	harina, harina para todo uso, harina de arroz, harina de repostería
10566	ajo	ajo, ajo picado, polvo de ajo McCormick R
10202	cebolla	cebolla, cebolla finamente picada, cebolla blanca, cebolla española picada
8681	agua	agua tibia (110 grados F), agua carbonatada

Figura 3.4: Número de ocurrencias, nombre y algunas descripciones diferentes para el diez ingredientes más frecuentes que se encuentran en las recetas de allrecipes.com.

## 3.2 Clasificación de dificultad de recetas

Cocinar es un proceso que implica muchas tareas y habilidades diferentes. Desde una receta describe tal proceso, puede ser visto como un algoritmo para producir algunos elaborados alimento. Los humanos pueden distinguir fácilmente entre diferentes niveles de complejidad de recetas. Desafortunadamente, las computadoras no tienen esa ventaja. Considere por ejemplo preparando unos "Sándwiches de pepino" frente a cocinar un "Pavo al romero Asar". Ya desde el nombre de la receta y sin ni siquiera mirar la receta descripción, mucha gente diría que el asado es más difícil que el sándwiches debido a sus experiencias pasadas.

### 3.2.1 Desarrollo de un método objetivo

Nuestro objetivo es encontrar una medida objetiva para expresar la dificultad de una receta, que no depende de nuestra percepción personal, sino sólo de la receta misma. Pero ¿Qué datos de receta debemos considerar? Bueno, una receta que requiere muchos diferentes las tareas en su preparación pueden ser más complejas que otra que requiere solo unos pocos pasos, por lo que elegimos considerar las instrucciones para determinar la receta dificultad. El tiempo necesario para preparar la receta también nos puede proporcionar alguna información útil sobre la dificultad. Así consideramos la preparación y tiempo total también. Por último, tenemos en cuenta el número de ingredientes que se requieren para la receta, porque cuantos más ingredientes se necesitan, mayor el esfuerzo por recogerlos y por organizar el entorno de trabajo. Suponemos que la dificultad depende linealmente de los parámetros elegidos y consideramos tres combinaciones de parámetros que se muestran en la Figura 3.1 como características para nuestra clasificación.

	Número de conjuntos	Número de puntos	de Preparación tiempo ingredientes	Tiempo total	Número de
Método 1	X			X	X
Método 2	X				X
Método 3			X		X

Tabla 3.1: Parámetros considerados en los diferentes métodos. Los conjuntos se refieren a elementos del conjunto {".", ",", ":", ";", "y", "o"}. Tanto el número de conectores como el de puntos se cuentan en la descripción de la receta. El tiempo se mide en minutos.

### 3.2.2 Evaluación de la validez de nuestros métodos

Como verdad básica, tomamos las recetas de [chefkoch.de](http://chefkoch.de) que contienen niveles de dificultad asignados por el usuario y usamos los métodos de clasificación Naive Bayes y Decision Trees. para determinar qué tan bien estas características etiquetan el nivel de dificultad. Para ambos métodos, los conjuntos de entrenamiento se generaron aleatoriamente tomando el 90% de los recetas, mientras que el 10% restante se utilizó como conjunto de prueba. Además diferentes Se usaron semillas para el generador aleatorio para construir diferentes conjuntos de entrenamiento.

## 3.3 Descubriendo similitudes de recetas

Durante siglos, diferentes culturas han desarrollado diferentes cocinas. con aumento globalización el intercambio entre diferentes cocinas ha crecido rápidamente y en consecuencia tenemos la posibilidad de comer alimentos y comprar ingredientes que eran previamente completamente desconocido para nosotros. Pero lo que creemos saber realmente corresponden a la realidad? ¿Y hay similitudes o diferencias que podamos no sabe y tal vez nunca imagine que sea posible? En esta sección pretendemos para investigar las diferencias y similitudes entre diferentes tipos de recetas. Nosotros por lo tanto, aplique dos algoritmos de aprendizaje automático diferentes a los datos de recetas recuperados del sitio web [allrecipes.com](http://allrecipes.com).

### 3.3.1 Similitud de recetas utilizando la agrupación en clústeres de K-Means

De manera similar a lo que se hizo en [6], definimos cada receta como un punto en k-dimensional Espacio euclidiano, cada dimensión representando un ingrediente. Nosotros consideramos solo los ingredientes más frecuentes a lo largo de nuestro conjunto de datos en lugar de todos los ingredientes y nosotros amplió la representación para considerar también la cantidad de ingrediente (por ejemplo ml o gramos) por ración, en lugar de considerar únicamente la presencia o ausencia de el ingrediente de la receta. Luego, cada receta se representa como un vector, donde el valor en cada índice de un ingrediente y se establece en 0 si el ingrediente no está presente.

Para visualizar mejor los datos utilizamos un esquema de reducción de dimensionalidad, que es particularmente adecuado para incrustar datos de alta dimensión en dos o tres dimensiones. Para evaluar visualmente la calidad del agrupamiento, coloreamos el puntos según la cocina o la categoría a la que pertenecen. La cocina y la información de la categoría se toma de allrecipes.com y se muestra en la Figura 3.5.

Categoría	Macro-cocina	Cocinas
Aperitivo	africano	Norteafricano / Sudafricano
Desayuno	asiático	chino/filipino/indio
Postre		Japonés / Coreano / Tailandés / Vietnamita
Ensalada	canadiense	
Pan de molde	européo	Europa del Este / Francés / Alemán / Griego / Italiano
Salsa y Condimento		Portugués / Escandinavo / Español / Inglés e Irlandés
Guarnición	Caribe latinoamericano / Mexicano / Sudamericano	
Sopa	Oriente Medio	
Beber	A NOBROTOS	Amish / Cajun / Judío / Nueva Inglaterra / Soul Food
Pollo		Meridional
Plato principal	Australia y Nueva Zelanda	

Figura 3.5: Lista de diferentes categorías, macrococinas y tipos de cocina.

### 3.3.2 Similitud de recetas usando árboles de clasificación

Con este enfoque, nuestro objetivo es encontrar el mejor árbol que muestre qué ingredientes son más relevantes al decidir si una receta está en una categoría o cocina específica. Por lo tanto, consideramos nuevamente el espacio definido por los ingredientes más frecuentes y construimos un árbol de clasificación usando todas las recetas. Los nodos internos de dicho árbol incrustará una condición binaria en un solo ingrediente, por ejemplo carne  $\leq 0,5$ , lo que significa que si se cumple la condición tomamos una rama y de lo contrario tomamos el otro. Esto refleja si un ingrediente está presente o no para el caso no ponderado y también en qué cantidad para el caso ponderado. Cada hoja incluye todas las recetas que satisfacen todas las condiciones a lo largo del camino desde el raíz hasta la hoja misma. Mirando la estructura final del árbol queremos ver qué ingredientes están más típicamente presentes en un subconjunto de recetas pero no en otros, y por lo tanto se puede utilizar para distinguir tales recetas.



# Implementación

---

En este capítulo pretendemos explicar las herramientas utilizadas a lo largo de este proyecto, así como dar una visión general de los scripts implementados para el análisis.

Cada script se implementó usando Python. Para rastrear los diferentes sitios web, usamos los módulos de solicitudes [\[9\]](#) para solicitudes HTTP simples y selenium [\[10\]](#) cuando teníamos que interactuar programáticamente con el navegador web porque se tenía que ejecutar algún código JavaScript en la página web de destino. Para analizar y buscar a través del código HTML descargado, usamos BeautifulSoup 4, que proporciona formas idiomáticas de navegar, buscar y modificar el árbol de análisis [\[11\]](#). Como Python 2.7 todavía tiene la codificación ASCII predeterminada y Unicode es la codificación más frecuente que se usa en Internet, incurrimos en algunos errores de decodificación al analizar el código HTML descargado. Para "reparar" el código roto usamos algunas funcionalidades provistas por `ftfy` [\[12\]](#).

Para extraer información de una cadena, usamos ampliamente las funcionalidades de expresiones regulares proporcionadas por el módulo `re`.

Para la clasificación de árboles, el agrupamiento de recetas y la clasificación de dificultad, usamos algoritmos provistos por el módulo `scikit-learn` [\[13\]](#), mientras que `numpy` [\[14\]](#) se usó para construir y manejar matrices.

Para la base de datos preferimos MongoDB [\[14\]](#) sobre MySQL [\[15\]](#) porque nos permitía mantener un esquema de base de datos realmente flexible.

## Capítulo 5

## Resultados

En este capítulo enumeramos nuestros resultados de la clasificación de dificultad de receta realizada, y al final, las observaciones cualitativas sobre la categoría y las similitudes culinarias son presentado.

### 5.1 Resultados de clasificación de dificultad

La Figura 5.1 muestra los valores medios medidos durante el proceso explicado en Sección 3.2 con 4 semillas generadoras aleatorias diferentes. En general tanto promedio preci

	Precisión		Recuerdo	
	árbol bayesiano	árbol bayesiano	árbol bayesiano	árbol bayesiano
método 1	0,66	0,66	método 2	0,67 0,65
0,65 0,64	método 3	0,60 0,65	0,65 0,63	
			0,60 0,63	

Figura 5.1: Precisión promedio y valores de recuperación para los diferentes métodos y algoritmos

El sion y el recuerdo promedio están por encima del 60%, lo que no está nada mal para un primer intento. con algoritmos simples. Observamos que la diferencia entre utilizar la clasificación Tree y Gaussian Naive Bayes es mínima, siendo como máximo del 5%. Con en la mayoría del 2% de diferencia, lo mismo se aplica a los dos primeros métodos, lo que significa que el tiempo de preparación y el tiempo total tienen una influencia similar en la dificultad de la receta. Sin embargo, el tercer método funciona un poco peor que los demás, por lo que podemos concluir que al analizar con mayor precisión las direcciones podemos mejorar la precisión un poco.

Sin embargo, si observamos los valores únicos que se muestran en las Figuras 5.2 y 5.3, tenemos una imagen diferente: es claramente visible que el primer nivel de dificultad tiene la mejor puntuación y que el porcentaje disminuye a medida que aumenta la dificultad. También es interesante que la subestimación es mucho más frecuente que la sobreestimación y que, en su mayor parte, todas las coincidencias incorrectas se etiquetan incorrectamente en el nivel de dificultad solo

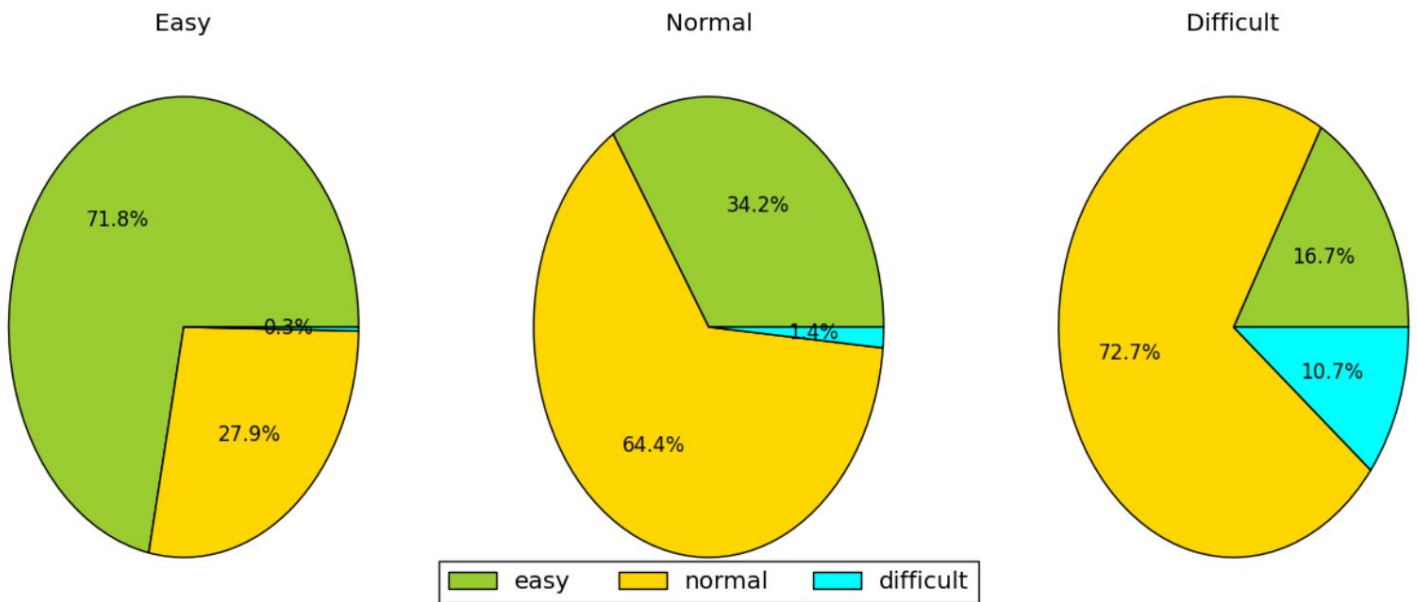


Figura 5.2: Gráficos circulares con valores de recuperación para cada nivel de dificultad de una ejecución de clasificación utilizando Clasificación de árboles con el método 1.

uno por debajo del esperado. Además, hay una diferencia más significativa entre los dos algoritmos que en el caso promedio, con respecto a la precisión de los dos primeros niveles: la clasificación de árbol tiene menor recuperación para recetas fáciles que Naive Bayes, pero compensa con un porcentaje más alto para el normal recetas.

Hay muchos factores que podrían haber influido en estos resultados y que podrían ser tomados en consideración para mejorar la clasificación. En primer lugar, las etiquetas de dificultad las proporciona el usuario que publicó la receta, lo que significa que son subjetivas y tal vez sesgadas. Considerando que es más probable que compartamos recetas que nos gustan o con las que estamos familiarizados, y que las personas tienden a subestimar la dificultad de una tarea o a sobrestimar sus habilidades (Efecto Dunning-Kruger [16]), es necesario asumir que el conjunto de datos está sesgado. Esto podría ser una explicación al curioso problema de subestimación de un nivel que notamos antes y también al bajo número de recetas etiquetadas como difíciles.

## 5. Resultados

decimals

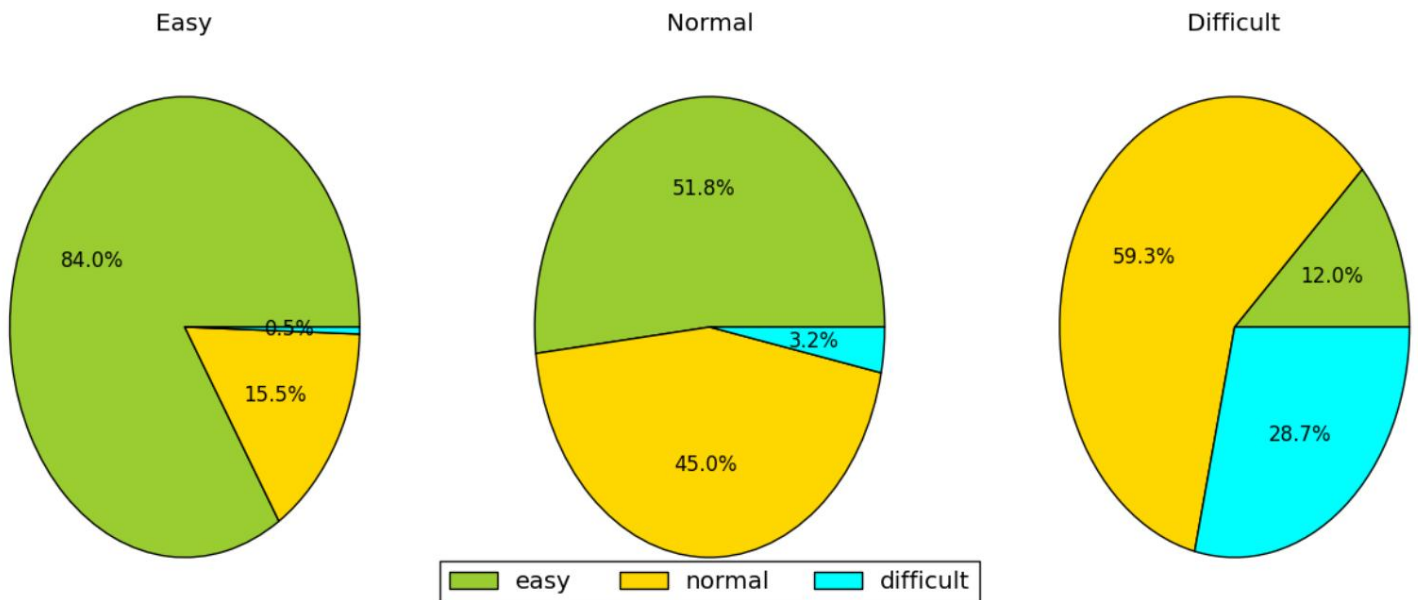


Figura 5.3: Gráficos circulares con valores de recuperación para cada nivel de dificultad de una ejecución de clasificación utilizando la clasificación Gaussian Naive Bayes con el método 1.

## 5.2 Similitudes de categoría

Es una tarea simple para los humanos separar los alimentos en diferentes categorías. Por ejemplo, los aperitivos, platos principales, guarniciones, etc. se comen en diferentes contextos y en diferentes momentos del día, por lo que pueden tener una estructura diferente.

Nuestro objetivo es encontrar similitudes y diferencias entre recetas de diferentes categorías observando solo sus ingredientes.

### 5.2.1 Agrupación de K-Means

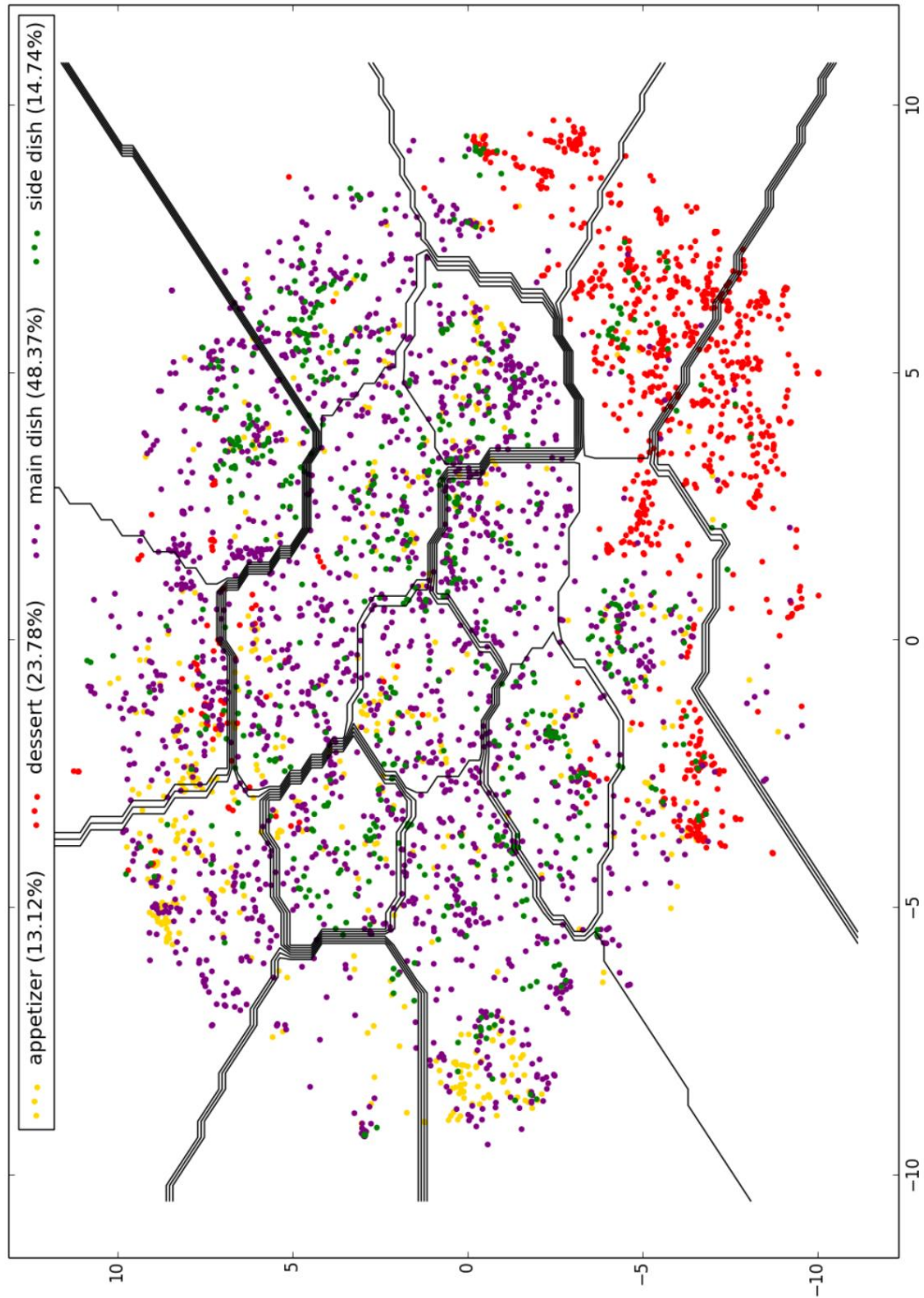
Para comenzar consideramos las siguientes 4 categorías:

- aperitivo
- guarnición
- plato principal
- postre

Esperaríamos ver grandes diferencias entre los postres y las otras tres categorías. También consideraríamos que los aperitivos se destacan un poco de las guarniciones y los platos principales, mientras que esperamos que las guarniciones y los platos principales compartan la mayoría de las características. Las Figuras 5.4 y 5.5 muestran la agrupación de todas las recetas pertenecientes a estas cuatro categorías, donde cada categoría se representa con un color diferente. En la primera figura usamos un gráfico de dispersión y en la segunda un gráfico circular para cada conglomerado. La figura 5.6 muestra los tres ingredientes más frecuentes de cada grupo y el gráfico circular correspondiente.

## 5. Resultados

18



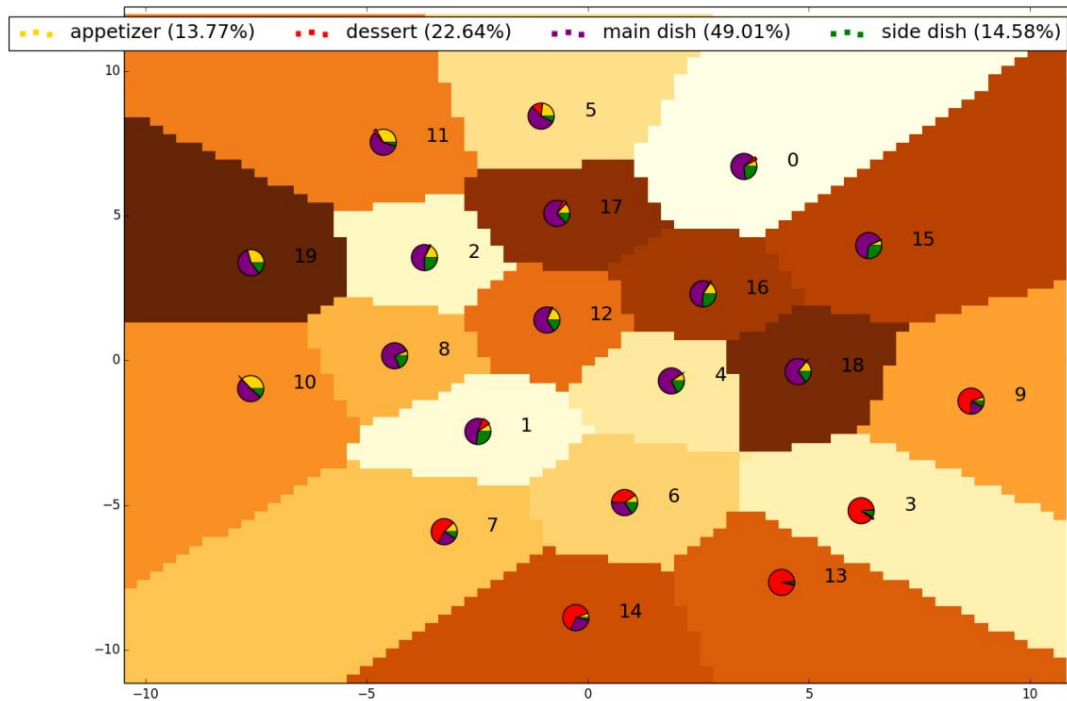


Figura 5.5: Agrupación de K-Means de las categorías de aperitivo, postre, plato principal y guarnición con 20 agrupaciones. Cada grupo está numerado y contiene un gráfico circular con la distribución de recetas contenidas entre las cuatro categorías.

La primera observación que podemos hacer es que los postres en realidad están bien separados de los otros platos porque en su mayoría cubren un área muy distinta del espacio. Los platos principales y secundarios están más distribuidos y cubren un área más amplia sin grandes concentraciones en áreas específicas. Los aperitivos se mezclan con platos principales y guarniciones, pero se concentran más en dos áreas más alejadas de los postres. Finalmente, no hay separación visible entre platos principales y guarniciones. Estas observaciones son aún más claras al observar la representación del pastel, donde también podemos ver que los ingredientes más frecuentes en los postres son el azúcar blanco, la mantequilla y la harina. En contraste con los platos mayores encontramos más a menudo especias (como pimienta negra y sal), ajo, cebolla, queso y aceite. Curiosamente, tenemos una imagen bastante diferente si consideramos también la cantidad de ingredientes. Como se puede ver en la Figura 5.7, la clara distinción entre los postres y las otras categorías ya no es visible, aunque podemos ver en la Figura 5.8 que los ingredientes más frecuentes para los postres son en su mayoría los mismos. Esto no es cierto para el resto de recetas, donde podemos notar novedades más frecuentes.

Diagrama de tarta de racimo	Ingredientes más frecuentes
0	mantequilla (68%) / pimienta negra (35%) / queso parmesano (34%)
1	agua (97%) / sal (54%) / cebolla (33%)
2	aceite vegetal (73%) / cebolla (57%) / pimienta negra (44%)
3	azúcar blanca (94%) / mantequilla (88%) / harina (61%)
4	ajo (94%) / aceite de oliva (87%) / sal (77%)
5	pollo (22%) / queso (15%) / nata (15%)
6	sal (69%) / harina (65%) / huevo (49%)
7	azúcar blanco (90%) / agua (60%) / sal (41%)
8	ajo (91%) / sal (82%) / cebolla (66%)
9	mantequilla (87%) / harina (61%) / sal (53%)
10	cilantro (91%) / sal (72%) / ajo (62%)
11	queso cheddar (58%) / crema agria (48%) / cebolla (32%)
12	sal (84%) / cebolla (48%) / pimienta negra (41%)
13	harina (96%) / azúcar blanco (89%) / sal (77%)
14	sal (94%) / harina (83%) / agua (83%)
15	mantequilla (97%) / sal (90%) / cebolla (48%)
disolvente	aceite de oliva (85%) / pimienta negra (66%) / sal (52%)
17	ajo (51%) / cebolla (38%) / salsa de soja (32%)
18	aceite de oliva (92%) / ajo (90%) / albahaca (52%)
19	sal (69%) / pimienta negra (54%) / ajo en polvo (52%)

Figura 5.6: Tres ingredientes más frecuentes y gráfico circular de todos los grupos de Figura 5.5.



ingredientes como carne de res y camarones. Además, en muchos clústeres ahora tenemos porcentajes más bajos para los ingredientes más frecuentes, lo que indica que el agrupamiento es peor.

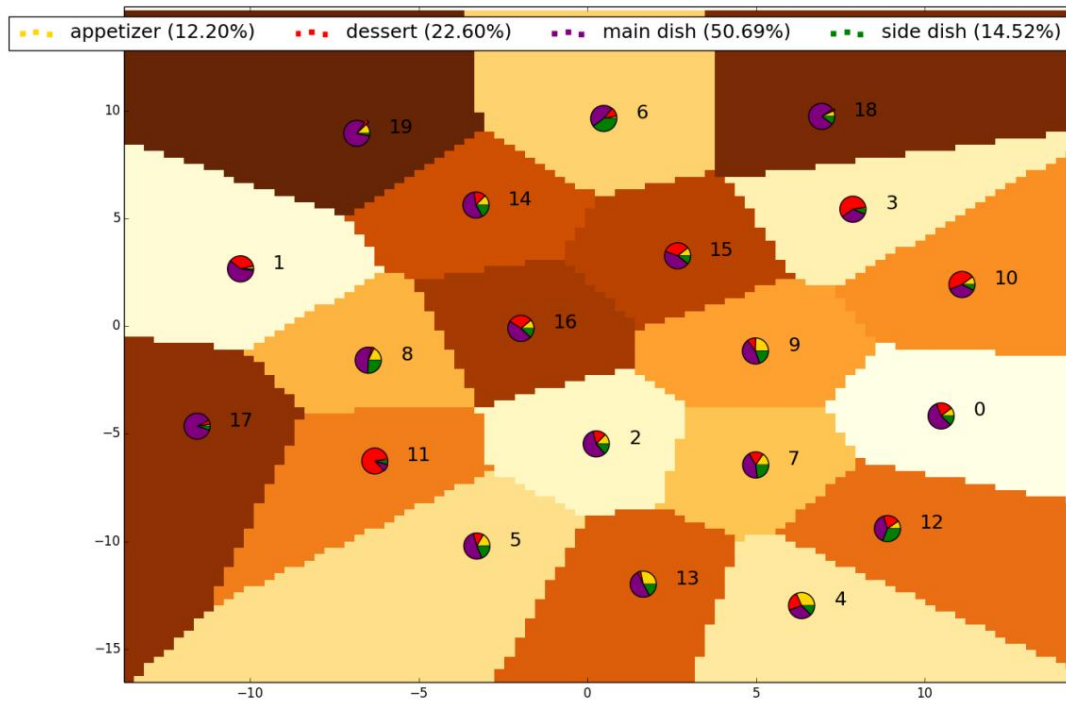


Figura 5.7: Agrupación de K-Means de las categorías de aperitivo, postre, plato principal y guarnición con 20 agrupaciones. También se considera el peso de los ingredientes. Cada grupo está numerado y contiene un gráfico circular con la distribución de recetas contenidas entre las cuatro categorías.

### 5.2.2 Clasificación de árboles

Nuevamente consideramos las categorías de la Subsección 5.2.1, pero esta vez usamos árboles de clasificación. La figura 5.9 muestra un árbol de profundidad 4 de recetas solo de las categorías de plato principal y postre del que se puede ver claramente que, como era de esperar, si una receta contiene azúcar, lo más probable es que sea un postre. Por otro lado, al observar los dos nodos internos que prueban la presencia de ajo, también vemos claramente que ambas hojas generadas que representan recetas que contienen ajo son puras y solo contienen platos principales.

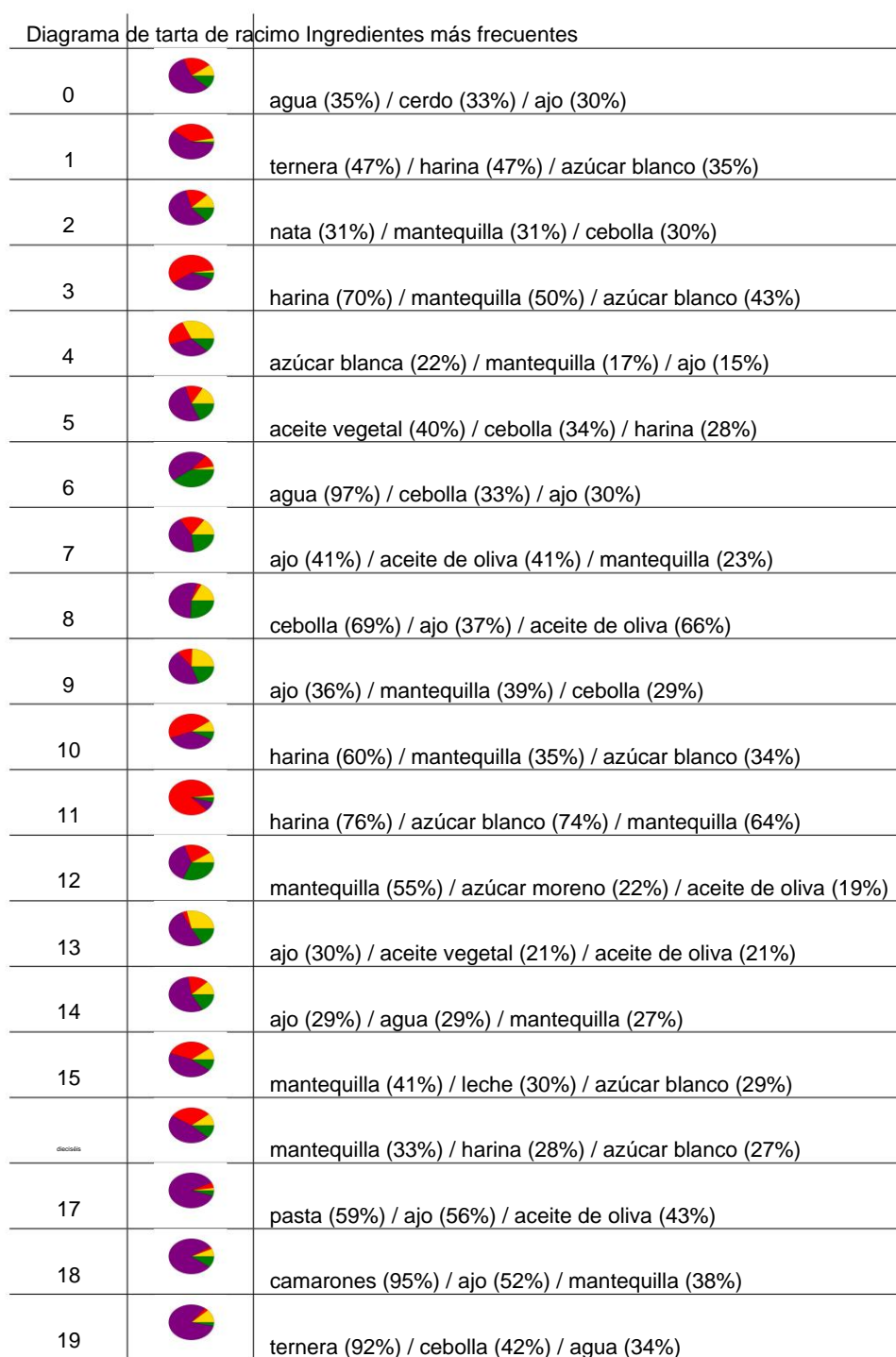
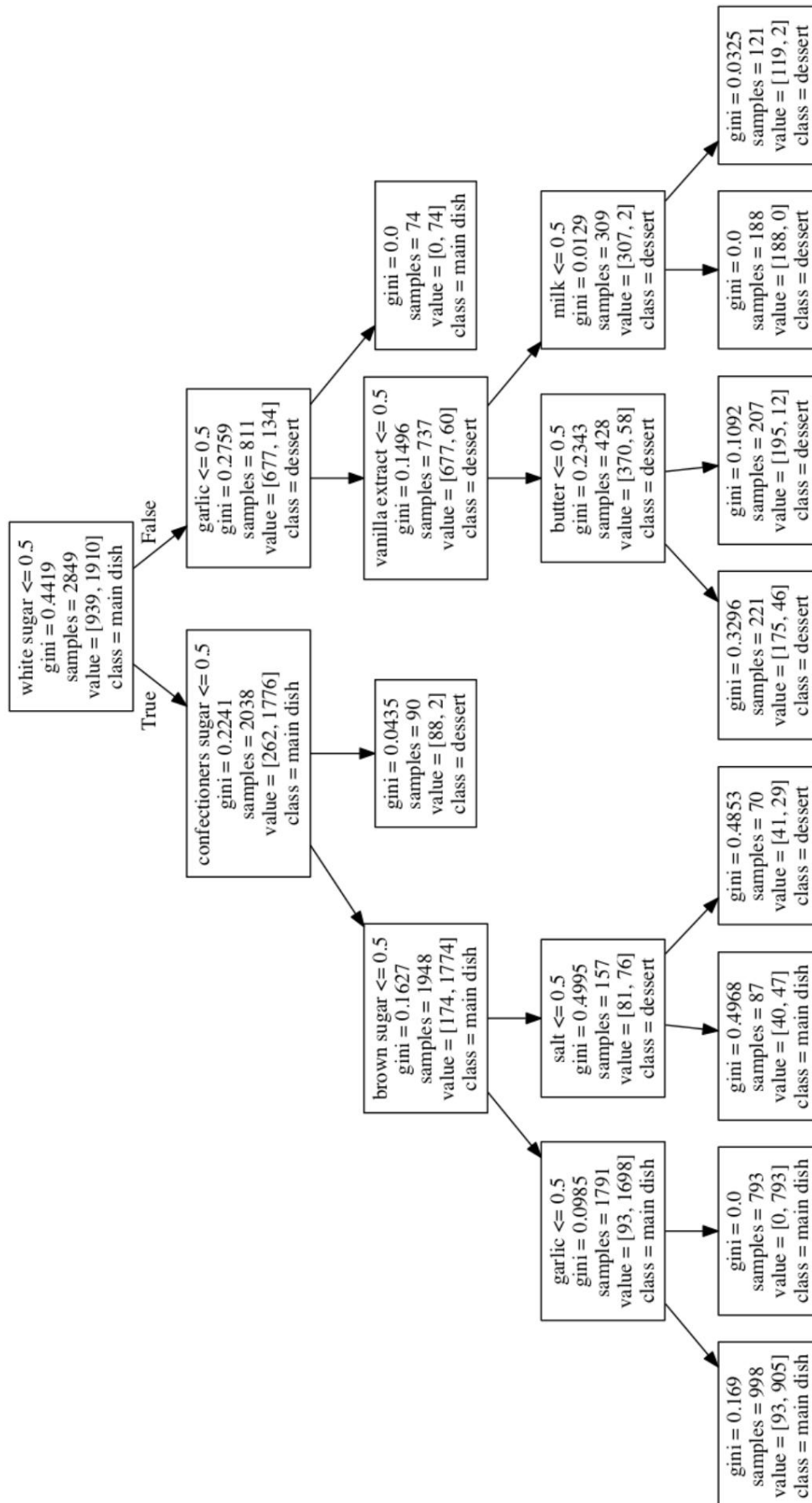


Figura 5.8: Tres ingredientes más frecuentes y gráfico circular de todos los grupos de Figura 5.7.

## 5. Resultados

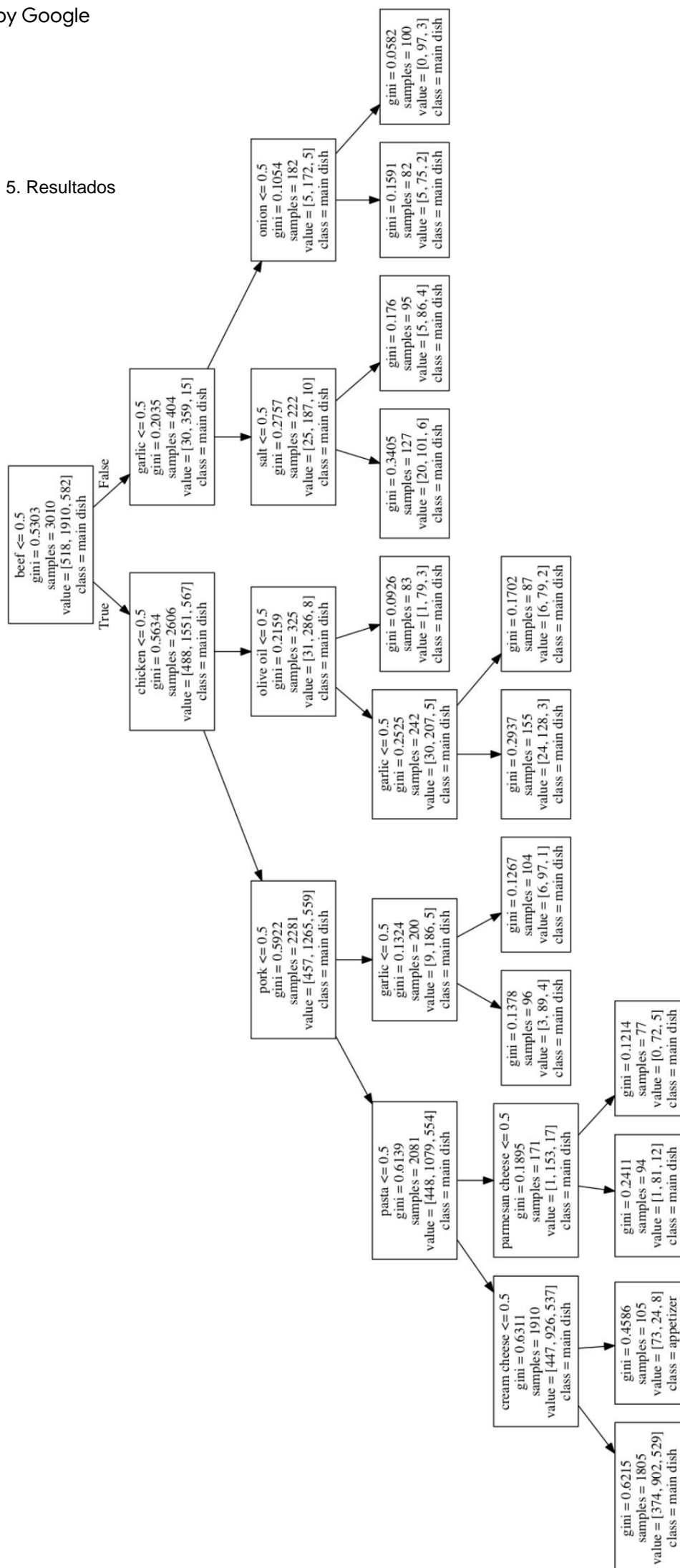
23



En la Figura 5.10 tenemos un árbol de profundidad 5 creado usando recetas de las categorías de aperitivos, platos principales y guarniciones. Aunque los números están bastante sesgados, todavía podemos ver que la presencia de carne o pasta es un indicador de los platos principales, mientras que el queso crema se encuentra con mayor frecuencia en los aperitivos, pero aparte de esto, no podemos ver ningún ingrediente que distinga una categoría. de los otros.

## 5. Resultados

25



## 5.3 Similitudes culinarias

La gente sabe un poco qué tipos de platos pueden esperar cuando van a un restaurante chino en lugar de a un restaurante italiano. Ahora usamos el mismo enfoque con las cocinas del mundo, para ver, por ejemplo, si es cierto que la cocina asiática es realmente diferente de la cocina europea.

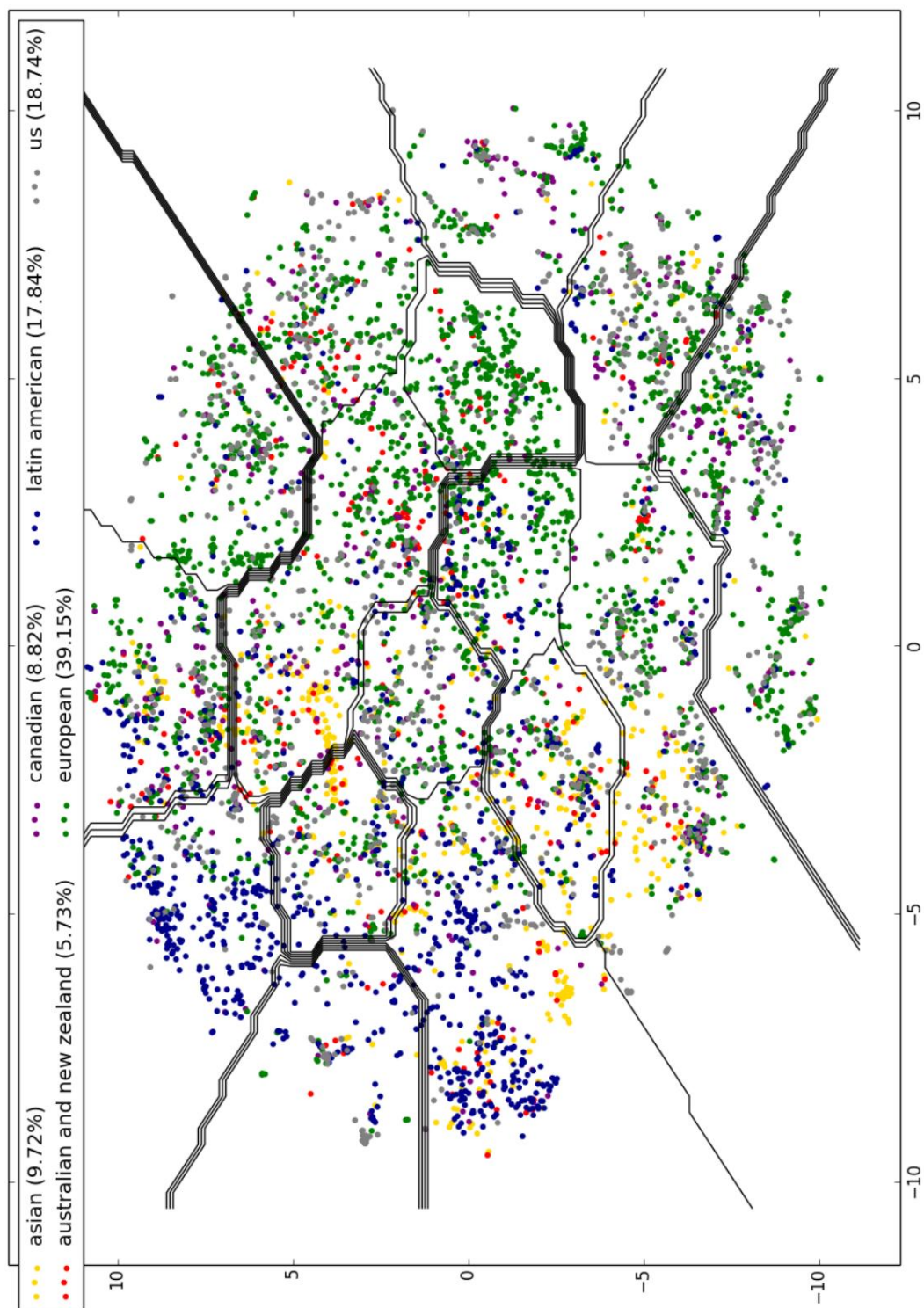
### 5.3.1 Agrupación de K-Means

Mirando las Figuras 5.11 y ??, que muestran la distribución de diferentes cocinas en el espacio de ingredientes, ya podemos hacer las siguientes observaciones:

- Las cocinas europea y latinoamericana están en su mayoría separadas;
- Las cocinas estadounidense, australiana y canadiense cubren todo el espacio;
- la cocina asiática se encuentra repartida entre las cocinas latinoamericana y europea.

## 5. Resultados

27



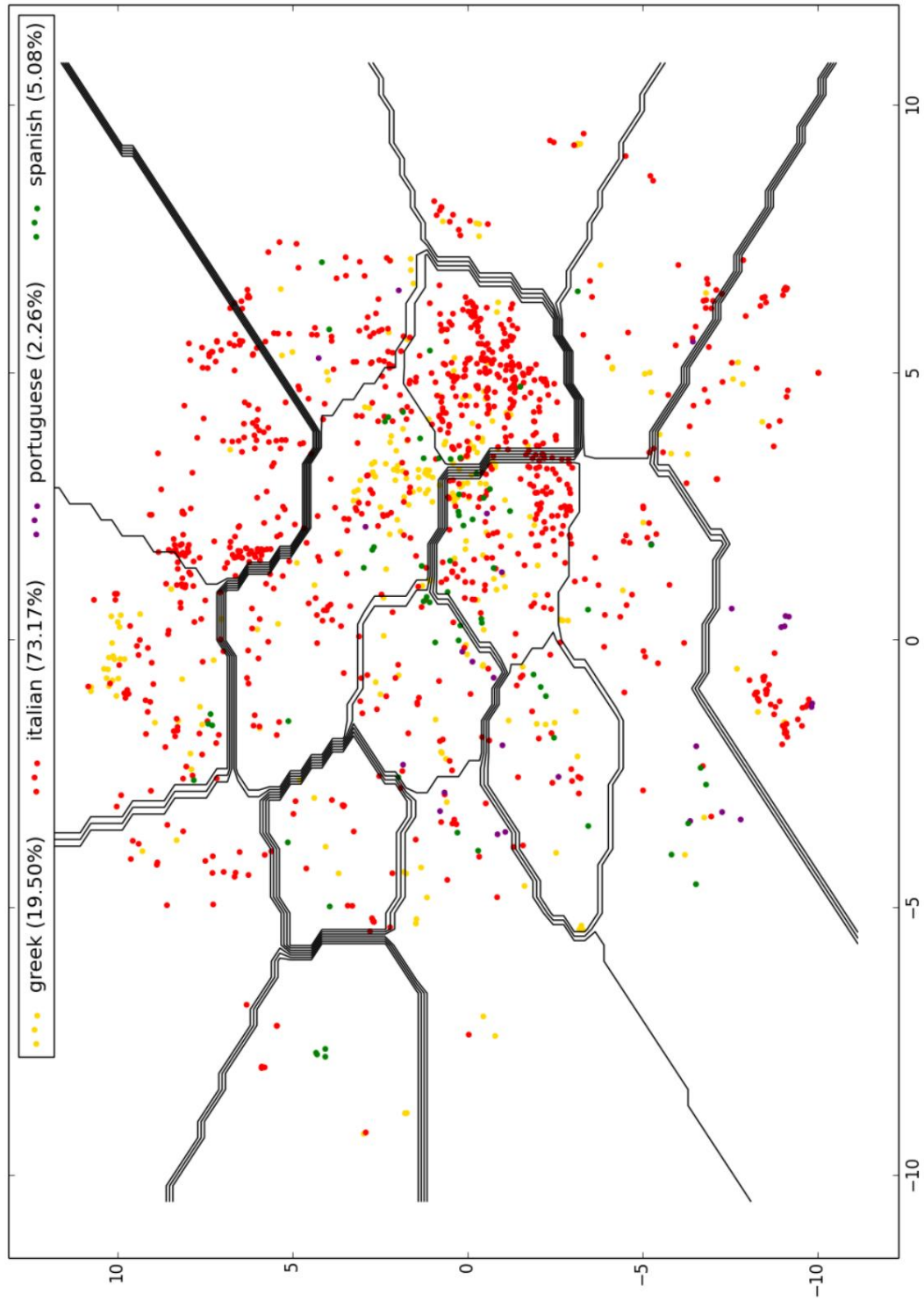
La separación de latinoamericanos y europeos es interesante un poco sorprendente, porque por razones históricas esperaríamos más conexiones, ya que los europeos difundieron su cultura en esas regiones y también trajeron nuevos ingredientes a Europa. Sin embargo, este fenómeno podría ser una explicación de las muchas similitudes entre las cocinas estadounidense y europea.

Dado que la macrococina europea no solo cubre una gran área en el espacio de los ingredientes, sino que también incluye muchas cocinas diferentes, como la italiana, la francesa, etc., queremos echar un vistazo más de cerca. Consideramos la conocida cocina mediterránea, que incluye las cocinas española, portuguesa, italiana y griega. Las Figuras 5.12 y 5.13 muestran su distribución en el espacio de ingredientes utilizando gráficos circulares y de dispersión.



## 5. Resultados

29



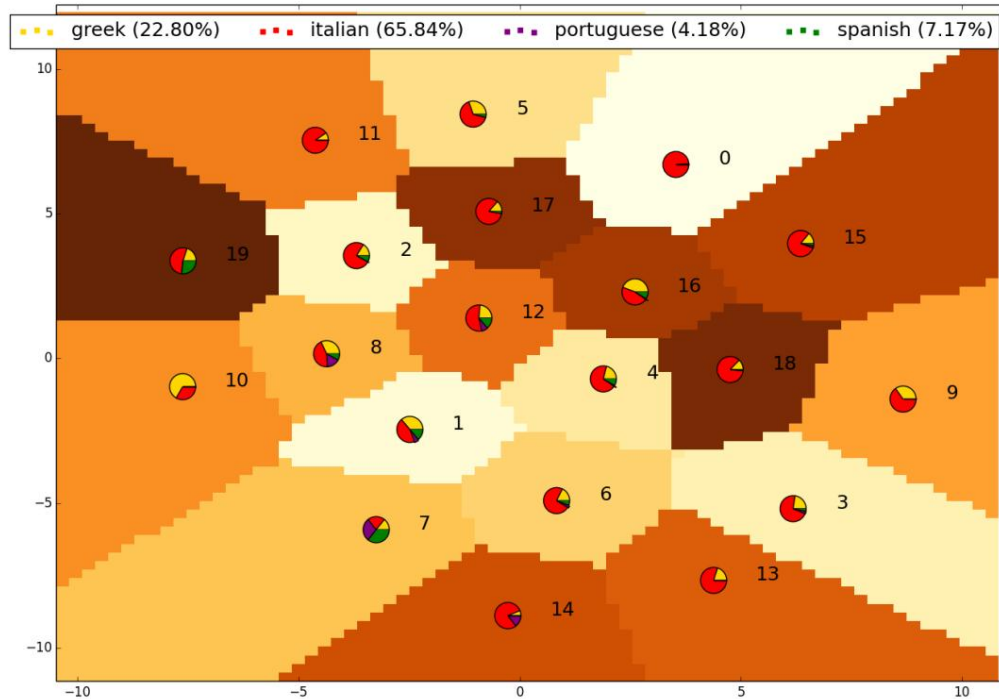


Figura 5.13: tarta mediterránea

Como era de esperar, es posible notar que hay muchas superposiciones, en particular entre las cocinas griega e italiana. Debido a que ocurren con menos frecuencia, es difícil saber si lo mismo ocurre con las cocinas española y portuguesa, pero podemos ver que no se concentran en una sola región.

Al observar la Figura 5.14, también es interesante ver que el grupo 0, que tiene Queso parmesano, mozzarella y mantequilla como ingredientes más frecuentes, contiene para la gran mayoría de las recetas italianas. Muy cerca se encuentra la región (clusters 4, 16, 18) con la mayor concentración de recetas de todos menos los portugueses cocina. Es claramente visible en la Figura 5.14 que la mayoría de las recetas mapeadas allí contienen aceite de oliva, que es uno de los ingredientes característicos de la dieta mediterránea.

### 5.3.2 Clasificación de árboles

Aquí queremos indagar un poco más en la cocina mediterránea para ver si hay ingredientes que no se comparten y se pueden considerar típicos de uno


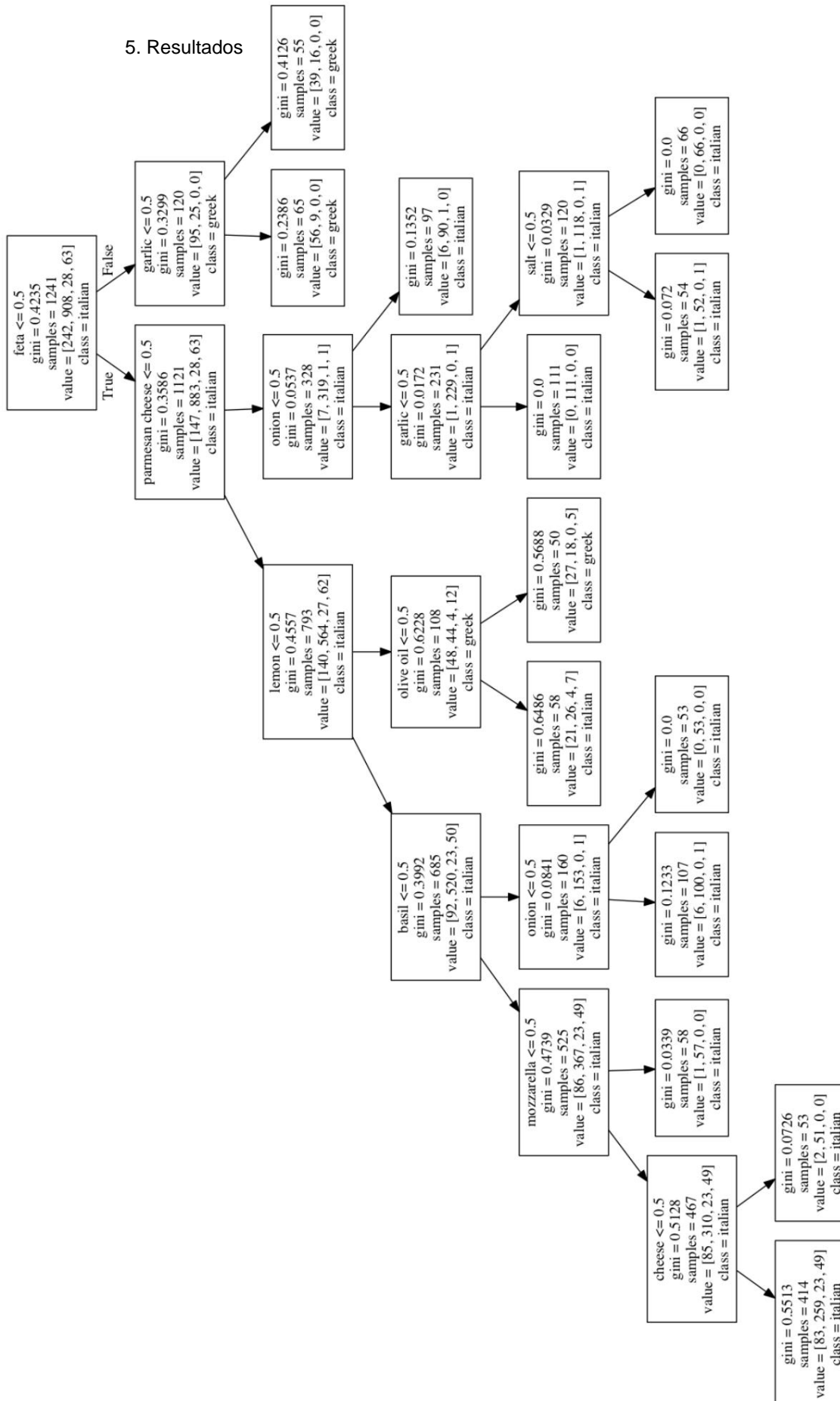
Diagrama de tarta de racimo	Ingredientes más frecuentes
0	 queso parmesano (63%) / mozzarella (50%) / mantequilla (47%)
1	 agua (86%) / sal (72%) / ajo (33%)
2	 cebolla (72%) / pimienta negra (60%) / aceite vegetal (56%)
3	 azúcar blanco (97%) / mantequilla (74%) / harina (68%)
4	 ajo (92%) / aceite de oliva (78%) / sal (71%)
5	 aderezo para ensalada (31%) / pollo (30%) / mozzarella (28%)
6	 huevo (59%) / harina (56%) / sal (50%)
7	 azúcar blanco (100%) / agua (50%) / limón (36%)
8	 listón (92%) / ajo (84%) / cebolla (56%)
9	 harina (96%) / mantequilla (91%) / sal (48%)
10	 ajo (100%) / sal (100%) / cilantro (100%)
11	 queso cheddar (61%) / ternera (48%) / cebolla (38%)
12	 sal (75%) / cebolla (55%) / ajo (37%)
13	 azúcar blanca (95%) / harina (93%) / polvo de hornear (65%)
14	 sal (96%) / levadura (96%) / agua (93%)
15	 mantequilla (93%) / sal (86%) / pimienta negra (52%)
disco	 aceite de oliva (83%) / pimienta negra (78%) / sal (51%)
17	 ajo (59%) / ternera (43%) / cebolla (38%)
18	 aceite de oliva (92%) / ajo (86%) / albahaca (60%)
19	 limón (47%) / perejil (40%) / albahaca (40%)

Figura 5.14: Tres ingredientes más frecuentes y gráfico circular de todos los grupos de Figura 5.13.

o más de las cocinas. Al observar la Figura 5.15 , vemos que el queso feta es un ingrediente presente principalmente en las recetas griegas, como era de esperar, ya que es originario de Grecia. Pero también hay una cierta cantidad de recetas italianas que lo contienen, lo que podría significar que esas recetas están mal etiquetadas como recetas italianas a pesar de ser recetas griegas, o que en algún momento hubo un intercambio entre las cocinas. También vemos que el queso parmesano juega un papel importante en las recetas italianas. Otros ingredientes claramente típicos solo de la cocina italiana son la albahaca, la mozzarella y el queso (que incluye muchas cualidades menos conocidas).

## 5. Resultados



## Perspectiva y Resumen

---

En esta tesis recopilamos un gran número de recetas y desarrollamos una gran base de datos de recetas estructuradas. Además, propusimos algunos modelos posibles para determinar la dificultad de una receta a partir de su descripción y utilizamos algoritmos de clasificación Gaussian Naive Bayes y Decision Tree para clasificar las recetas en diferentes categorías de dificultad. Además, utilizamos el agrupamiento de K-Means y los árboles de clasificación para buscar similitudes y diferencias en el espacio de ingredientes entre recetas de diferentes categorías y cocinas del mundo. Un posible trabajo futuro podría ser mejorar la base de datos de recetas para obtener una estructura uniforme independientemente de la fuente y el idioma de los datos, así como incluir todos los ingredientes en el análisis y no solo los más frecuentes.

Otro posible trabajo futuro podría ser recopilar más información, como el costo y el valor nutricional, para ingredientes separados, lo que nos permitiría desarrollar una aplicación de sugerencias de recetas saludables para usuarios que desean ayuda para comer de manera más saludable. Para mejorar el análisis de dificultad, podríamos considerar más parámetros como la presencia de ingredientes específicos o las palabras separadas en los pasos de dirección de cocción. Esto podría hacerse con un análisis gramatical para contar el número de acciones necesarias y dar mayor peso a las acciones más involucradas.

Este trabajo ha sentado las bases para futuros estudios de recetas al generar una gran colección de datos estructurados de recetas.

# Bibliografía

- [1] : Allrecipes: página de inicio. <http://todaslasrecetas.com/> [Acceso: 2015-12-17].
- [2] : Epicurious: página de inicio. <http://www.epicurious.com/> [Acceso: 2015-12-17].
- [3] Svensson, M., Höök, K., Cöster, R.: Diseño y evaluación de kalas: A sistema de navegación social para recetas de comida. Transacciones de ACM sobre interacción humana entre computadoras (TOCHI) 12(3) (2005) 374–400
- [4] Ueda, M., Takahata, M., Nakajima, S.: extracción de preferencias alimentarias del usuario para recomendaciones personalizadas de recetas de cocina. Gestión Semántica de la Información Personalizada: Recuperación y Recomendación SPIM 2011 (2011) 98
- [5] Teng, CY, Lin, YR, Adamic, LA: recomendación de recetas usando redes de ingredientes. En: Actas de la 4ª edición anual de ACM Web Science Conferencia, ACM (2012) 298–307
- [6] Nedovic, V.: Aprendizaje del espacio de ingredientes de recetas usando la probabilidad generativa modelos En: Actas del Taller de Cocina con Computadoras (CwC). Volumen 1. (2013) 13–18
- [7] Su, H., Shan, MK, Lin, TW, Chang, J., Li, CT: Clasificación automática de recetas cui sine por ingredientes. En: Actas de la Conferencia Conjunta Internacional ACM de 2014 sobre Computación Ubicua y Pervasiva: Adjunto Publicación, ACM (2014) 565–570
- [8] Friedman, J., Hastie, T., Tibshirani, R.: Los elementos del aprendizaje estadístico. Volumen 1. Serie Springer en estadísticas Springer, Berlín (2001)
- [9] : Solicitudes: página de inicio. <http://docs.python-requests.org/en/latest/> [Acceso: 2015-12-17].
- [10] : Selenium: página de inicio. <http://www.seleniumhq.org/> [Acceso: 2015-12-17].
- [11]: BeautifulSoup: página de inicio. <http://www.crummy.com/software/BeautifulSoup/> [Acceso: 2015-12-17].
- [12] : Ftfy: Repositorio Github. <https://github.com/LuminosoInsight/python-ftfy> [Acceso: 2015-12-17].

- [13]: Scikit-learn: página de inicio. <http://scikit-learn.org/stable/> [Accedido: 2015-12-17].
- [14]: MongoDB: página de inicio. <https://www.mongodb.org/> [Acceso: 2015-12-17].
- [15] : MySQL: página de inicio. <https://www.mysql.com/> [Acceso: 2015-12-17].
- [16] Dunning, D., Johnson, K., Ehrlinger, J., Kruger, J.: Por qué la gente no reconoce su propia incompetencia. *Direcciones actuales en la ciencia psicológica* 12(3) (2003) 83–87
- [17] : Datos de autoalimentación: página de inicio. <http://nutritiondata.self.com/> [Acceso: 2015-12-17].



## Apéndice Capítulo

---

Medida de cocción	Peso (gramos)
taza	125
cucharadita	4
cucharada	12
clavo pizca	8
raya cubo	2
	5
	15
rodaja	12
racimo	3

Figura 7.1: Tabla de conversión para diferentes medidas no estándar que se encuentran en las descripciones de los ingredientes.

ingrediente	Peso (gramos)
huevo	53
(roja) cebolla	110
calabacín (roja)	196
patata lima	213
	67
tomate	123
{verde, rojo, amarillo} pimiento 119	
zanahoria	61
palta	201
champiñón bollo	18
de manzana	110
melocotón clara	140
de huevo yema	150
de huevo	29
	13
naranja	159
plátano	118
puerro	89
limón	108
Coco	397
pimiento rojo	905
piña	45
mozzarella	93
de salchicha	30
calabaza	1500
brócoli	500
pepino	301
repollo	35
lechuga	260
salmón de	6
judías verdes	159
espárragos	<small>desconocido</small>
atún con	207
mango	180
coliflor	216
alcachofa	128
tomate cherry	<small>desconocido</small>
corazón de alcachofa	100
pescado encurtido	57
	37
pera	178
muslo de pollo	140
albaricoque	35
berenjena tomate	458
seco	3

Figura 7.2: Tabla de conversión para la densidad de diferentes ingredientes (considerando tamaño mediano promedio de [17]).