



**ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL**  
**Facultad de Ingeniería en Electricidad y Computación**

“Métricas de colaboración para trabajo autónomo dentro de sistemas de gestión de aprendizaje”

**TRABAJO DE TITULACIÓN**

Previa a la obtención del Título de:  
**MAGISTER EN CIENCIAS DE LA COMPUTACIÓN**

**Presentado por:**  
Alberto Jiménez Macías

**GUAYAQUIL - ECUADOR**

**AÑO: 2018**

## **AGRADECIMIENTOS**

Agradezo a Dios darme las fuerzas para superar los obstáculos y dificultades a lo largo de este camino.

A Centro de Tecnologías de la Información(CTI) por darme las facilidades para realizar la maestría.

A mi tutor Dr. Sixto García por dedicación y apoyo en el presente trabajo.

## DEDICATORIA

Dedico este trabajo primero a Dios por permitirme llegar hasta este momento importante en mi formación profesional. A mis padres, por ser los pilares fundamentales en mi vida. A mi esposa, por su apoyo incondicional y motivación para esta culminación. A mis hermanos Estefani, Mario y Jorge . A mi sobrina Charlotte, por ser el ángel que llena de amor nuestras vidas. A mi abuelita Máxima , por sus consejos.

**TRIBUNAL DE SUSTENTACIÓN**

---

PhD. Sixto García

---

PhD.  
PROFESOR DELEGADO POR EL DECANO DE LA FACULTAD

## **DECLARACIÓN EXPRESA**

“La responsabilidad por los hechos, ideas y doctrinas expuestas en este informe me corresponde exclusivamente; y, el patrimonio intelectual de la misma, a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL”

(Reglamento de Graduación de la ESPOL).

---

Alberto Jiménez Macías

## RESUMEN

Algunos algoritmos de minería de datos se han utilizado para calcular contribución realizada por los participantes al crear un documento o wiki colaborativa. En el presente trabajo se propone un modelo utilizando un variante de Non-negative Matriz Factorization(NMF) para el cálculo de la contribución de cada participante. El modelo se divide en tres secciones: Obtención de revisiones(Data Access), Algoritmo de medición de colaboración (Collaboration Measurement Logic) y Visualizador Web (Web Visualization). Para el análisis se utilizó archivos de Google Drive enviados mediante el sistema de gestión de aprendizaje utilizado en ESPOL (SIDWEB) , los cuales fueron integrados al modelo propuesto mediante el Visualizador Web. El modelo se puede adaptar para mostrar la colaboración de Wikipedia, One Drive. Los resultados del algoritmo de medición de colaboración mediante un modelo adaptativo de NMF dieron los mejores resultados comparados contra otras métricas tales como: contador de palabras, contador de oraciones, contador de ediciones, similaridad mediante grafos, tópicos usando Latent Dirichlet Allocation(LDA). Se realizó un prototipo de Visualizador Web para poder integrar el modelo propuesto a cualquier Sistema de Gestión de Aprendizaje.

# ÍNDICE GENERAL

RESUMEN	VI
INDICE GENERAL	VII
INDICE DE TABLAS	IX
INDICE DE FIGURAS	X
ABREVIATURAS	X
<b>1. INTRODUCCIÓN</b>	<b>1</b>
1.1. Antecedentes y Justificación . . . . .	1
1.2. Objetivos . . . . .	2
1.2.1. Objetivo General . . . . .	2
1.2.2. Objetivos Específicos . . . . .	2
1.3. Alcance . . . . .	3
1.4. Organización del documento . . . . .	3
<b>2. TRABAJOS RELACIONADOS y METODOLOGÍA</b>	<b>4</b>
2.1. Algoritmos para medición de contribución . . . . .	5
2.2. Metodología . . . . .	10
<b>3. MÉTRICA Y MODELO</b>	<b>12</b>
3.1. Métrica . . . . .	12

3.2. Construcción y evaluación del modelo . . . . .	14
3.2.1. Obtención de revisiones (Data Access) . . . . .	15
3.2.2. Lógica de medición de colaboración (Collaboration Measurement Logic) . . . . .	16
3.2.3. Visualizador Web (Web Visualization) . . . . .	16
3.3. Colección de datos . . . . .	16
3.4. Herramientas de implementación . . . . .	17
<b>4. RESULTADOS y ANÁLISIS</b>	<b>18</b>
4.1. Lógica de medición de contribución . . . . .	18
4.1.1. Métrica de evaluación . . . . .	19
4.1.2. Resultado para la métrica propuesta . . . . .	19
4.2. Visualización y comparación con evaluación manual desde SIDWEB . . . . .	23
<b>5. CONCLUSIONES</b>	<b>26</b>
<b>BIBLIOGRAFÍA</b>	<b>28</b>

## ÍNDICE DE TABLAS

Tabla 4.1. Trabajo grupal calificado individualmente . . . . .	24
Tabla 4.2. Comparación entre Nota SIDWEB y Nota Manual usando prototipo . . . . .	24
Tabla 4.3. Grupo 1: Mayor contribución de 1 estudiante . . . . .	25
Tabla 4.4. Grupo 2: Contribución compartida entre 3 estudiantes . .	25

# ÍNDICE DE FIGURAS

3.1.	Contribución por usuario . . . . .	13
3.2.	Modelo propuesto . . . . .	15
4.1.	Análisis comparativo de precisión . . . . .	20
4.2.	Análisis del hiperparámetro $\beta$ . . . . .	21
4.3.	Comparación de la duración de los algoritmos . . . . .	22
4.4.	Nivel de contribución de los estudiantes en documento . . . . .	22
4.5.	Prototipo del Visualizador Web . . . . .	23

## ABREVIATURAS

**NMF** - Non-negative Matrix Factorization

**LDA** - Latent Dirichlet Allocation

# CAPÍTULO 1

## 1. Introducción

### 1.1. Antecedentes y Justificación

En las universidades de todo el mundo, los docentes pueden dedicar una cantidad significativa de tiempo a revisar tareas y proyectos grupales presentados por sus alumnos. Las tecnologías basadas en la web, como Google Docs, han proporcionado una plataforma para que los estudiantes escriban documentos en colaboración.

Actualmente, esas plataformas brindan información limitada sobre la contribución individual hecha por cada estudiante. Los estudios previos [1] [2] [3] [4] se han centrado en los aspectos cuantitativos de la contribución de los individuos en la escritura colaborativa, mientras que el aspecto de la calidad ha recibido menos atención.

Un modelo que permita medir la calidad de la colaboración

realizada por los estudiantes en un documento es importante para el docente porque permite obtener evidencias entre la interacción de los estudiantes y mejorar el proceso de enseñanza/aprendizaje.

## **1.2. Objetivos**

### **1.2.1. Objetivo General**

Proponer un modelo para la medición de la colaboración individual en trabajos colaborativos.

### **1.2.2. Objetivos Específicos**

Los objetivos específicos del presente trabajo son los siguientes:

1. Diseñar una metodología para la medición de colaboración en trabajos colaborativos .
2. Implementar la metodología usando diversas técnicas de minería de datos para la educación.
3. Evaluar el rendimiento de las técnicas al ser aplicadas al mismo conjunto de datos.
4. Comparar los resultados de rendimiento entre las diferentes técnicas.
5. Implementar e integrar un prototipo de visualización a un sistema de gestión de aprendizaje (SIDWEB).

6. Ayudar al profesor en proporcionarle información que le permita comprender y analizar la contribución de cada uno de los estudiantes .

### **1.3. Alcance**

El alcance del presente trabajo es el uso de métodos de minería de datos en la educación para proponer un modelo de medición de contribución en documentos colaborativos y desarrollo de un prototipo para poder integrarlo a un sistema de gestión de aprendizaje. Se considero trabajos enviados mediante Google Drive en la plataforma SIDWEB para los experimentos.

### **1.4. Organización del documento**

El presente documento se encuentra estructurado de la siguiente manera: En el Capítulo 1 se exponen los antecedentes y la justificación del presente trabajo, además se plantean los objetivos que se pretenden alcanzar. El Capítulo 2 se hace una revisión de los trabajos existentes relacionados a la medición de contribución en documentos o wikis de forma colaborativa. El Capítulo 3 se propone una metodología para la medición de contribución mediante el análisis de diversas técnicas de minería de datos para la educación. En el Capítulo 4 se exponen los resultados de la aplicación del modelo y se comparan los valores de rendimiento. Finalmente se presentan las conclusiones y recomendaciones del trabajo realizado.

## CAPÍTULO 2

# 2. Trabajos Relacionados y Metodología

Se han propuesto varios métodos para medir la contribución de los usuarios en la redacción colaborativa. La mayoría de las investigaciones anteriores se han centrado en Wikipedia o en Git<sup>1</sup> debido a su naturaleza de fuente abierta, mientras que pocos estudios se basan en Google Docs u Office 365. Todo los métodos propuestos son para documentos escritos en Inglés, existe poca información para Español. No obstante, todos estos sistemas realizan un seguimiento de cada revisión (por ejemplo, edición o acción de compromiso) en los documentos. Por lo tanto, se pueden utilizar los métodos propuestos

---

<sup>1</sup><https://git-scm.com/>

a la fuentes mencionadas para medir las contribuciones.

## 2.1. Algoritmos para medición de contribución

En el contexto del desarrollo de software con sistemas de administración de código fuente como Apache Subversión [5] o Git, la medición de contribución se ha analizado en términos de propiedad del código. La medición de la calidad del código fuente aún se basa en el seguimiento a nivel de línea [6]. Este seguimiento propuesto también se utiliza como una unidad básica para identificar contribuyentes. En este método, el seguimiento a nivel de línea permite la identificación del usuario que realizó el último cambio en una línea específica de código en un documento, pero la información sobre el creador original se pierde. Esta funcionalidad es apropiada para entornos de desarrollo de software colaborativo, ya que permite la detección del usuario responsable de introducir defectos o realizar cambios en el código. Sin embargo, este mecanismo no es adecuado para rastrear a los contribuyentes originales del contenido o para detectar cambios a un nivel más detallado, como palabras o caracteres.

En el contexto de Wikipedia, se han propuesto varios enfoques para medir las contribuciones de los usuarios. Viegas en [1] atribuye el contenido de una oración al usuario que realizó el último cambio. Sin embargo, no reconoce correctamente al autor del contenido reintroducido después de haber sido eliminado. Otro inconveniente de trabajar a nivel de oraciones es el hecho de que los pequeños cambios (por ejemplo, agregar palabras, corregir la gramática o deshacer cambios) no se identifican. Por lo tanto, este método no

mide correctamente la contribución general de los usuarios.

Ding en [7] desarrolló una visualización de wikis empresariales a gran escala, midiendo las contribuciones de los usuarios en función del número de ediciones realizadas por cada usuario en diferentes páginas. Hess en [8] también ha propuesto un método para medir el alcance de la contribución parcial de un usuario en cada versión del documento. Compara la versión actual con la anterior, y la contribución global es la suma de todas las contribuciones parciales. Basado en un enfoque similar, Sabel en [9], calculó las contribuciones de los usuarios a cada versión del documento y luego utilizó los resultados como la calificación de los usuarios en un sistema de reputación.

Korfiatis en [10] estimó las contribuciones de los usuarios a Wikipedia, basadas en el análisis de redes sociales, utilizando específicamente indicadores como la centralidad de los usuarios como un proxy de su reputación. Hoisl en [11] implementó un complemento para plataformas wiki que permite la medición de la contribución parcial de los usuarios según las diferencias entre las versiones del documento. Luego, asigna un peso a cada contribución parcial en función de una métrica de importancia (por ejemplo, número de vistas o votos).

Estos estudios previos han utilizado diferentes enfoques para resolver el problema de medir las contribuciones en la escritura colaborativa. Sin embargo, estos enfoques tienen varias limitaciones:

1. El método de medición básico se basa en un simple conteo de ediciones en un documento [7], por lo tanto, puede ser fácilmente manipulado por los usuarios que buscan aumentar

su puntaje de contribución. Los enfoques que usan el análisis de redes sociales [10] son buenos para calcular la distribución de las contribuciones de los usuarios en múltiples documentos wiki, pero no proporcionan una buena estimación del nivel de contribución a un documento específico.

2. Los métodos anteriores no tienen en cuenta la diferencia entre las contribuciones que permanecen en el documento durante largos períodos de tiempo y las contribuciones que se eliminan rápidamente [9, 8]. Por ejemplo, una contribución que se elimina rápidamente recibe el mismo puntaje en comparación con una contribución que permaneció en un documento durante mucho tiempo. Es probable que las contribuciones de mayor calidad permanezcan a lo largo del tiempo; por lo tanto, los algoritmos deberían considerar la duración de una contribución. En el caso de Google Docs, aunque el lapso de tiempo de la contribución no es tan grande como las páginas Wiki, se debe considerar el mismo principio.
3. Además, los algoritmos para medir la contribución deben verificar el tipo de contribuciones que están capturando (por ejemplo, nuevo contenido, cambios de formato) y estimar el grado en que ese tipo de contribución representa la contribución general de los usuarios. Pfeil en [12] propuso una metodología para cerrar esta brecha mediante la utilización de una teoría fundamentada para categorizar los tipos de contribución en wikis, que luego fue utilizada por Ehmann en [13]. Esta categorización tiene en cuenta las contribuciones más allá de agregar nuevo conteni-

do, por ejemplo, deshacer información existente, correcciones gramaticales, eliminar información irrelevante o incluso aclarar información.

Basado en [12, 13], Arazy en [4] propuso una categorización más simple, al considerar tipos específicos de contribución que pueden detectarse mediante algoritmos, como agregar nuevo contenido, deshacer cambios, vincular interno o externo, eliminar contenido y corregir. En su algoritmo, la unidad básica de significado es una oración. Esto se basa en la noción de propiedad de la oración, en la que los usuarios poseen las oraciones que crearon. La propiedad de una oración continúa si más del 50 % de las palabras son las mismas entre dos revisiones consecutivas del documento. Las métricas obtenidas representan el número de oraciones que un usuario ha agregado o eliminado. Además, el algoritmo considera la cantidad de enlaces, así como los cambios a nivel de palabra.

Adler en [14] propuso *Wikitrust*, un etiquetado visual de secciones confiables y no confiables en un documento wiki. El algoritmo encuentra las coincidencias más largas para todas las secuencias de palabras entre dos revisiones consecutivas, pero elimina fragmentos de palabras. Por lo tanto, puede detectar palabras reintroducidas y rastrear la autoría del contenido, ya que realizar retrocesos a versiones anteriores es una práctica común en Wikipedia. Sin embargo, el algoritmo utiliza un enfoque codicioso, y generalmente cae en óptimas locales. Esto puede conducir a interpretaciones erróneas de la autoría de las palabras, especialmente cuando las secuencias de palabras se mueven en lugar de ser eliminadas o insertadas.

De Alfaro en [2] introdujo un algoritmo para medir las contribu-

ciones de los usuarios que requiere comparar cada nueva revisión con todo el conjunto de revisiones anteriores, teniendo en cuenta que el contenido se puede eliminar y luego volver a insertar. Para cada sección de contenido en la última revisión, el algoritmo busca todas las versiones anteriores si hay secciones similares de contenido que estadísticamente es poco probable que sucedan naturalmente, estableciendo así la autoría adecuada. Usando estas coincidencias, determina la autoría más antigua de cada token (por ejemplo, palabras) en el nuevo contenido. El tiempo de ejecución del algoritmo es relativo a la agregación del tamaño de la última revisión y el número total de ediciones en el historial de revisión. Sin embargo, el algoritmo no fue probado en términos de precisión.

Flóck en [3] presentó un enfoque de modelo de árbol para establecer la autoría de contenido y, por lo tanto, mide la contribución de los usuarios en un documento. En esta versión inicial, el modelo solo consideró párrafos y oraciones; produjo una precisión de aproximadamente 60 %, lo que no es adecuado para su uso. Un algoritmo mejorado [15], basado en un modelo de gráfico k-partite, construye una representación de nivel fino del documento al considerar párrafos, oraciones y tokens. Este modelo fue más eficiente e informó un aumento del 30 % en la precisión.

La métrica planteada mide la contribución realizada por el estudiante durante la construcción del documento tomando en cuenta la calidad de la aportación.

## 2.2. Metodología

Los documentos enviados en Google Drive mediante SIDWEB fueron evaluados manualmente por personas para obtener la contribución realizada por cada estudiante. Con los mismos documentos se obtuvo la contribución de forma automática utilizando las siguientes métricas:

1. **Conteo Edición (EC)** es un conteo del número de ediciones de los estudiantes en un documento [7].
2. **Conteo de palabras (WC)** es un recuento del número de palabras de cada estudiante aportado en un documento [15].
3. **Conteo palabras claves (WCK)** es similar al método WC, pero excluye *stop words* y extrae la raíz de las palabras restantes.
4. **Similaridad grafos (GS)** consiste en representar cada documento como un gráfico [16]. En el gráfico del documento, cada palabra representa un nodo y los enlaces se establecen entre palabras consecutivas. Luego, creamos un gráfico para cada estudiante basado en las palabras de su autoría. La contribución del alumno es la sub-gráfica común máxima entre el documento y los gráficos del alumno. La representación gráfica captura la estructura del documento, pero el inconveniente es que los nodos aislados en los gráficos de los estudiantes no se resumen en su contribución general.
5. **Palabras clave de similitud de grafos (GSK)** es similar al método GS, y realiza limpieza de datos como en la métrica WCK.

6. **Conteo oraciones(SC)** es un recuento del número de oraciones que posee cada usuario en el documento. La propiedad de las oraciones se basa en el modelo propuesto por Arazy en [4].
7. **Tópicos NMF (NMF)** es nuestro modelo propuesto que utiliza NMF para extraer temas del documento.
8. **Tópicos LDA (LDA)** amplía nuestro modelo, pero utiliza la asignación latente Dirichlet (LDA) en lugar de NMF para extraer temas de los documentos.

La métrica planteada obtuvo mejores resultados con respecto a la evaluación manual realizada. Se construyó un modelo para poder evaluar la métrica dentro de un sistema de gestión de aprendizaje y poder brindar una solución al docente al momento de calificar un trabajo colaborativo enviado a sus estudiantes.

## **CAPÍTULO 3**

### **3. Métrica y Modelo**

A continuación se describe el proceso planteado para calcular la contribución realizada por cada estudiante dentro del documento utilizando la métrica planteada. Para integrarlo con los sistemas de gestión de aprendizaje se diseñó un modelo y pueda ser utilizado por el docente en el proceso de evaluación de un trabajo colaborativo.

#### **3.1. Métrica**

El presente estudio propone los pasos indicados en la figura 3.1 para medir la contribución realizada por cada estudiante en un documento.

A continuación se explica con más detalle cada uno de los pasos a seguir:

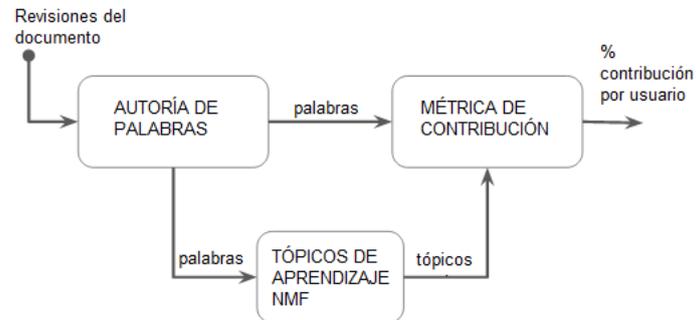


Figura 3.1: Contribución por usuario

1. **Autoría de palabras (Tokens Authorship):** Se establece la autoría de las palabras de un documento y se amplía el modelo basado en grafos propuesto por Flock en [15]. Este modelo se implementó específicamente para páginas de Wikipedia, pero se puede generalizar a documentos de Google u otros sistemas de escritura colaborativa que utilizan un control de versionamiento.
2. **Tópicos de aprendizaje NMF (NMF Topics Learning):** Se asocia la autoría de ideas principales con un alto nivel de contribución al documento. Para inferir las ideas principales aplicamos el modelado de temas. Para extraer los temas  $t_i$  del contenido del documento, usamos una factorización matricial adaptativa no negativa (NMF). Se eligió modelado de temas basado en NMF porque ofrece mejores resultados, como se muestra en la sección 4.
3. **Métrica de contribución(Contribution Measurement):** Para medir la contribución general de los estudiantes en documentos colaborativos, la métrica planteada considera las métricas tanto

cuantitativas como cualitativas. La medición cuantitativa se basa en la cantidad de palabras que cada alumno contribuyó al documento. La medida cualitativa considera en qué medida el estudiante contribuye a los temas del documento. La contribución de un estudiante se define de la siguiente manera:

$$C_s = \beta * Tc_s + (1 - \beta)Wc_s \quad (3.1)$$

Donde  $Tc_s$  se refiere a la cantidad de palabras asociadas a los temas que forman parte de la colección de nodos que pertenecen a cada estudiante:

$$Tc_s = \sum_{i=1}^n tw_i \quad \forall tw \in G(T) \quad (3.2)$$

La medida cuantitativa es un conteo simple de las palabras contribuidas por cada estudiante al documento.

$$Wc_s = \sum_{i=1}^n t_i \quad \forall t \in G(T) \quad (3.3)$$

### 3.2. Construcción y evaluación del modelo

El modelo diseñado puede ser utilizado por cualquier Sistema de Gestión de Aprendizaje (Learning Management System) mediante un Visualizador Web, con el objetivo de ayudar al docente en la calificación de un documento colaborativo. En la medición del nivel de

contribución de los estudiantes se considera tanto la cantidad como la calidad del aporte realizado. Para la evaluación del modelo se utilizó como fuente de datos los archivos de Google Docs de los trabajos enviados por los estudiantes a través del SIDWEB<sup>1</sup>.

En la figura 3.2 se explica el modelo propuesto, indicando la forma de interacción entre las partes, SIDWEB representa al Sistema de Gestión de Aprendizaje.

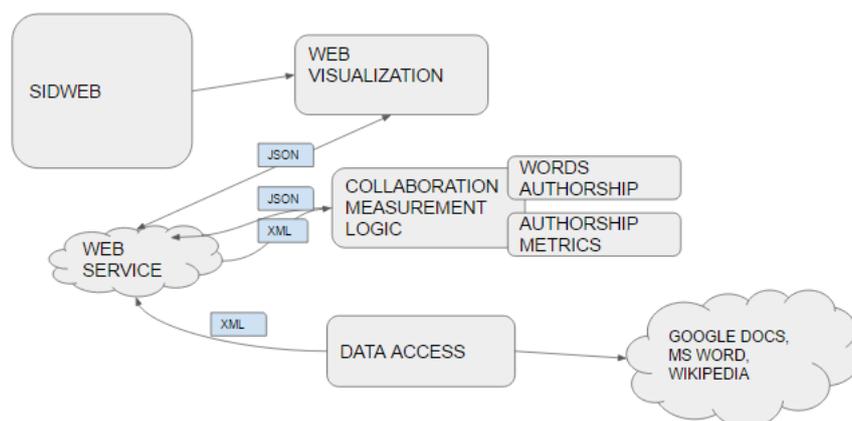


Figura 3.2: Modelo propuesto

A continuación se procede a describir cada sección del modelo:

### 3.2.1. Obtención de revisiones (Data Access)

Para obtener las revisiones del documento se utilizó la función Revisions del Api de Google Drive y se generó un archivo XML con la información de las revisiones de los documentos. Se puede adaptar a cualquier servicio que permita escritura colaborativa por ejemplo: Google Drive, One Drive, Wikipedia.

<sup>1</sup>LMS utilizado por ESPOL

### **3.2.2. Lógica de medición de colaboración (Collaboration Measurement Logic)**

Con la información del archivo XML generado en el paso anterior se procedió a calcular la contribución por usuario como lo indica en la sección 3.1 . Esta sección retorna en formato JSON la información correspondiente a los participantes del documento incluyendo la contribución realizada por cada uno.

### **3.2.3. Visualizador Web (Web Visualization)**

El visualizador web recibe en formato JSON la información proveniente del paso anterior. Permite a cualquier sistema obtener de forma visual los resultados del algoritmo propuesto. Se realizará cambios en la presentación de la información del sistema con la finalidad de dar una mejor experiencia a los docentes que utilicen esta herramienta para calificar sus trabajos colaborativos.

## **3.3. Colección de datos**

La información utilizada consta de 39 documentos de Google Drive con un total de 763 revisiones, realizadas por 92 estudiantes . Los documentos fueron enviados a un sistema de aprendizaje de ESPOLE por estudiantes como parte de sus tareas o proyectos grupales. Para la recopilación de las revisiones de los documentos se utiliza el API de Google Drive.

### **3.4. Herramientas de implementación**

El modelo esta implementado en Python para las partes de Obtención de revisiones y Lógica de medición de colaboración utilizando algoritmos de minería de datos. El Web Service también esta realizado en Python. El prototipo del Visualizador Web está escrito en Python , HTML, JQuery y CSS

# **CAPÍTULO 4**

## **4. Resultados.**

A continuación se detallan los resultados de los experimentos realizados con los documentos de Google Drive enviados desde SIDWEB:

### **4.1. Lógica de medición de contribución**

Para evaluar la precisión de nuestra métrica al medir la contribución de los estudiantes, se compara con otros métodos de referencia. En la siguiente subsección, la métrica de evaluación y los resultados obtenidos.

#### 4.1.1. Métrica de evaluación

Para cada documento, calculamos el error absoluto medio (MAE) dado por:

$$\text{MAE}_{\text{doc}} = \frac{1}{n} \sum_{i=1}^n |f_i - y_i|. \quad (4.1)$$

Donde  $f_i$  es la contribución del usuario calculado y  $y_i$  es nuestro valor de verdad del terreno determinado manualmente por los revisores humanos.

#### 4.1.2. Resultado para la métrica propuesta

Se comparó los métodos basados en el MAE acumulativo para la contribución de los estudiantes en cada documento. La figura 4.1 muestra que los modelos basados en temas (NMF, LDA) superan a todos los demás métodos, de los cuales NMF tiene un MAE menor. El tercer MAE inferior corresponde al método WC, mientras que EC es el peor.

En la métrica propuesta, se ha evaluado el hiperparámetro  $\beta$  (*beta*) para determinar qué valor produce una mejor precisión. La figura 4.2 muestra el MAE acumulativo para diferentes valores de  $\beta$  en el modelo NMF. Los valores más altos de *beta* producen un error menor, y encontramos que la mejor configuración es 0,9.

En relación con la escalabilidad, los modelos basados en temas son más lentos en comparación con los métodos simples basados en recuentos. Esto se debe a la complejidad algorítmica de los modelos NMF y LDA. La complejidad del

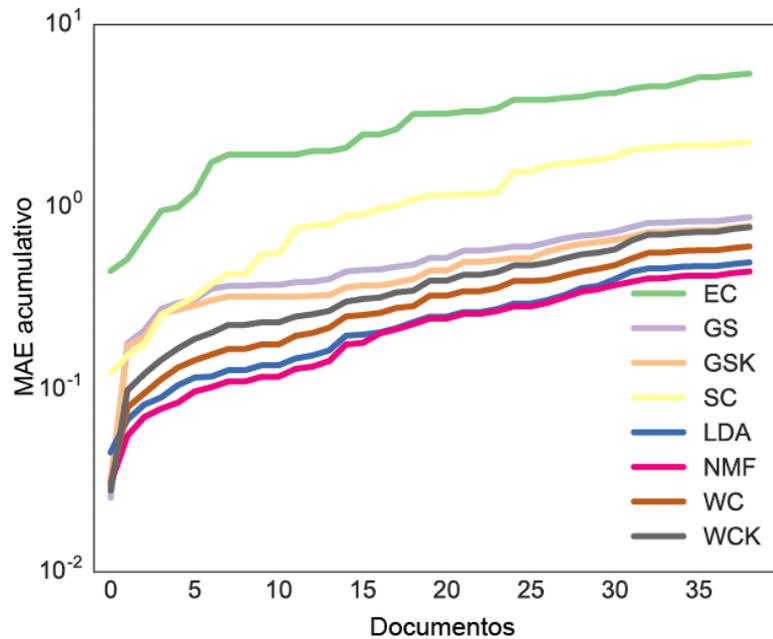


Figura 4.1: Análisis comparativo de precisión

tiempo es polinomial en NMF [17]. En LDA, la complejidad temporal es proporcional al número de muestras e iteraciones [18]. La figura 4.3 muestra claramente que hay tres grupos de algoritmos de acuerdo con la complejidad del tiempo. Los métodos más lentos y complejos, en la sección superior del gráfico, son NMF, LDA, GSK y EC. En el caso de EC, aunque es un método de conteo simple, debe iterar todas las revisiones de los documentos, lo cual es un proceso que consume tiempo. El siguiente grupo, con respecto a la complejidad del tiempo, incluye GS y WCK. Y, en la sección inferior del gráfico, los algoritmos más rápidos son los métodos simples de conteo WC y SC.

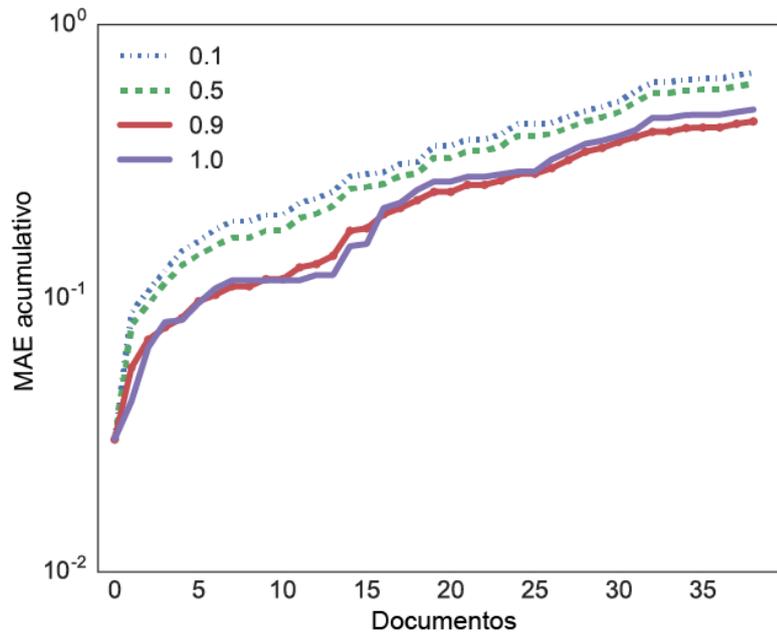


Figura 4.2: Análisis del hiperparámetro  $\beta$ .

Aunque los modelos basados en temas son más lentos que los métodos de referencia, la duración depende del tamaño del documento y del número de revisiones. Por lo tanto, el tiempo de espera para las aplicaciones que usan nuestro modelo será relativo a estos dos factores [15]. En nuestro conjunto de datos, para documentos con un tamaño de  $7K$  palabras, la duración oscila entre 10 y 13 segundos.

En otros aspectos de los resultados, la métrica permite detectar a aquellos estudiantes que hacen una contribución baja en los documentos. La figura 4.4 muestra que el 69% de estudiantes aportaron 10% o menos a los documentos.

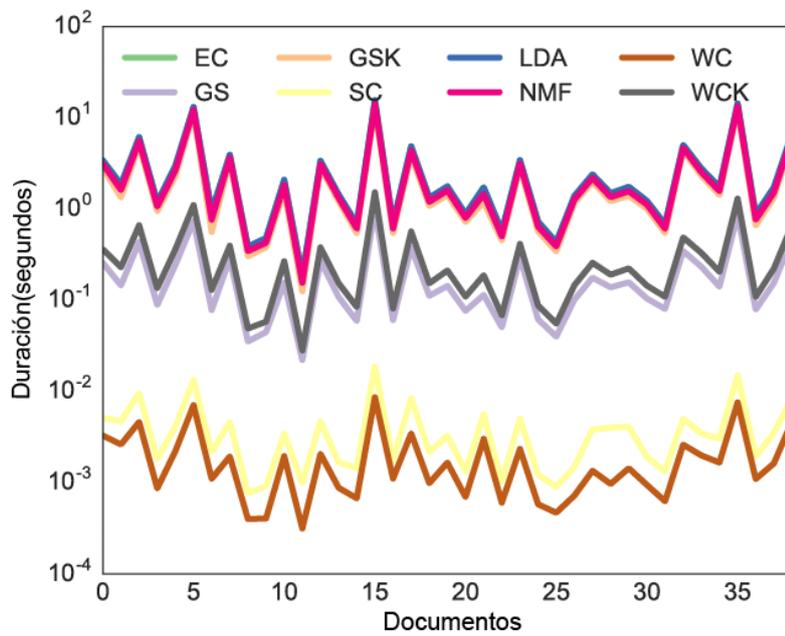


Figura 4.3: Comparación de la duración de los algoritmos

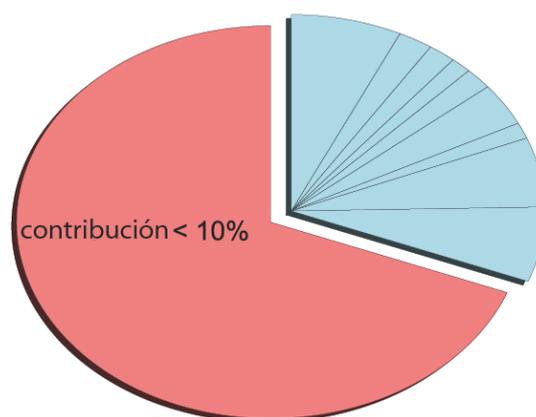


Figura 4.4: Nivel de contribución de los estudiantes en documento

## 4.2. Visualización y comparación con evaluación manual desde SIDWEB

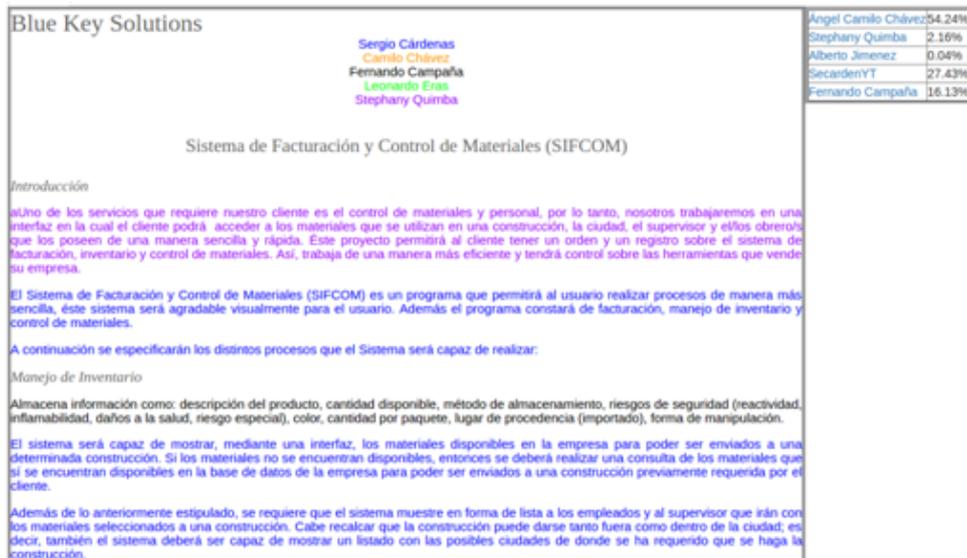


Figura 4.5: Prototipo del Visualizador Web

El Visualizador Web que se muestra en la figura 4.5 es un prototipo funcional desarrollado en Python con la finalidad de evaluar el modelo planteado dentro de SIDWEB. Se observa la versión final del documento y el listado de los estudiantes que contribuyeron. La columna final indica el porcentaje de colaboración realizada por cada estudiante en el documento calculado por medio de la métrica planteada en el presente estudio, para poder observar el aporte realizado por cada estudiante se debe dar click sobre el mismo.

Se utilizó el prototipo en SIDWEB con 39 tareas enviadas por los estudiantes por medio de Google Drive. Los resultados obtenidos usando el modelo son comparados con la evaluación manual realizada por el docente. A continuación se muestra tres muestras de los

resultados obtenidos :

<b>Estudiantes</b>	<b>Métrica Contribución</b>
Estudiante 1	85.75 %
Estudiante 2	12.62 %
Estudiante 3	1.63 %

Tabla 4.1: Trabajo grupal calificado individualmente

En la tabla 4.1 se muestra el resultado de un trabajo grupal con puntaje máximo de 40 puntos realizado en el II término del 2016 de la materia XYZ. El grupo estaba compuesto por cuatro estudiantes. Podemos observar los resultados usando el modelo propuesto que solo tres estudiantes contribuyeron en la escritura del documento. Los resultados nos demuestran que el Estudiante1 realizó una contribución casi nula.

<b>Estudiantes</b>	<b>Nota SIDWEB</b>	<b>Nota Usando Prototipo</b>
Estudiante 1	35	40
Estudiante 2	30	10
Estudiante 3	25	0
Estudiante 4	25	0

Tabla 4.2: Comparación entre Nota SIDWEB y Nota Manual usando prototipo

Se procedió a realizar una evaluación manual por un docente usando el prototipo con la información de las contribuciones realizadas por cada estudiante. La tabla 4.2 muestra las notas obtenidas mediante SIDWEB y la nota realizada por el docente usando la herramienta planteada. Podemos observar la diferencia entre las notas pero la segunda tiene una mayor relación con la contribución realizada por el estudiante. El modelo propuesto permite que el docente visualice la contribución de cada estudiante de un grupo,

dándole la opción de poder calificar en base a la aportación del estudiante y no solo en base al documento final.

<b>Estudiantes</b>	<b>Métrica Contribución</b>
Estudiante 1	62.14 %
Estudiante 2	20.95 %
Estudiante 3	11.41 %
Estudiante 4	5.45 %
Estudiante 5	0.05 %

Tabla 4.3: Grupo 1: Mayor contribución de 1 estudiante

<b>Estudiantes</b>	<b>Métrica Contribución</b>
Estudiante 1	35.91 %
Estudiante 2	32.85 %
Estudiante 3	31.24 %

Tabla 4.4: Grupo 2: Contribución compartida entre 3 estudiantes

En las tablas 4.3 y 4.4 podemos observar el desempeño de tres estudiantes en dos diferentes grupos. Esta información es útil para el docente porque permite agrupar a los estudiantes dependiendo del nivel de contribución realizada. Esto permite que los estudiantes que tengan bajo nivel de contribución puedan desarrollar el resultado de aprendizaje de trabajo en equipo en su formación profesional.

## **CAPÍTULO 5**

### **CONCLUSIONES**

El enfoque presentado se centran en el aspecto cuantitativo de las contribuciones individuales, pero éste no es el único factor que los instructores evalúan en la escritura colaborativa. El aspecto cualitativo es importante, porque algunos estudiantes pueden escribir menos que sus compañeros, pero pueden contribuir con contenido más significativo.

El modelo propuesto ayuda al docente en la formación de la habilidad en los estudiantes para trabajar como parte de un equipo. La herramienta permite a los docentes identificar problemas en el desempeño de los estudiante durante cualquier instante de tiempo que dure la escritura colaborativa y realizar los correctivos necesarios con la finalidad de mejorar la calidad del documento. El modelo propuesto proporciona ideas útiles sobre el proceso de edición para

la escritura colaborativa.

Los resultados obtenidos por cada estudiante permite identificar las habilidades de cada uno, con esta información podríamos crear nuevos grupos con mejor distribución de estudiantes. Según el análisis de los resultados el tamaño del grupo de trabajo fluye en la contribución realizada de manera inversamente proporcional, se debería poner un límite en el número de integrantes de cada grupo.

Como trabajo futuro se considera realizar cambios en la presentación de la información del Visualizador Web con la finalidad de dar una mejor experiencia de usuario a los docentes que utilicen esta herramienta para calificar sus trabajos colaborativos. También se recomienda compartir los documentos de Google Drive a un único correo para obtener mayor información al utilizar el API de Google Drive.

# BIBLIOGRAFÍA

- [1] Fernanda B Viégas, Martin Wattenberg, and Kushal Dave. Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 575–582. ACM, 2004.
- [2] Luca De Alfaro and Michael Shavlovsky. Attributing authorship of revisioned content. In *Proceedings of the 22nd international conference on World Wide Web*, pages 343–354. ACM, 2013.
- [3] Fabian Flöck and Andriy Rodchenko. Whose article is it anyway?—detecting authorship distribution in wikipedia articles over time with wikigini. *Online proceedings of the Wikipedia Academy*, 2012.
- [4] Ofer Arazy and Eleni Stroulia. A utility for estimating the relative contributions of wiki authors. In *ICWSM*, 2009.
- [5] C Michael Pilato, Ben Collins-Sussman, and Brian W Fitzpatrick. *Version control with subversion*. "O'Reilly Media, Inc.", 2008.
- [6] Robert Baggen, José Pedro Correia, Katrin Schill, and Joost Visser. Standardized code quality benchmarking for improving software maintainability. *Software Quality Journal*, 20(2):287–307, 2012.

- [7] Xianghua Ding, Catalina Danis, Thomas Erickson, and Wendy A Kellogg. Visualizing an enterprise wiki. In *CHI'07 extended abstracts on Human factors in computing systems*, pages 2189–2194. ACM, 2007.
- [8] Michael Hess, Brian Kerr, and Lev Rickards. Wiki user statistics for regulating behaviour. Technical report, Citeseer, 2006.
- [9] Mikalai Sabel. Structuring wiki revision history. In *Proceedings of the 2007 international symposium on Wikis*, pages 125–130. ACM, 2007.
- [10] Miguel-Angel Sicilia, Nikolaos Th Korfiatis, Marios Poulos, and George Bokos. Evaluating authoritative sources using social networks: an insight from wikipedia. *Online Information Review*, 30(3):252–262, 2006.
- [11] Bernhard Hoisl, Wolfgang Aigner, and Silvia Miksch. Social rewarding in wiki systems—motivating the community. In *International Conference on Online Communities and Social Computing*, pages 362–371. Springer, 2007.
- [12] Ulrike Pfeil, Panayiotis Zaphiris, and Chee Siang Ang. Cultural differences in collaborative authoring of wikipedia. *Journal of Computer-Mediated Communication*, 12(1):88–113, 2006.
- [13] Katherine Ehmman, Andrew Large, and Jamshid Beheshti. Collaboration in context: Comparing article evolution among subject disciplines in wikipedia. *First Monday*, 13(10), 2008.
- [14] B Thomas Adler, Luca de Alfaro, Ian Pye, and Vishwanath Raman. Measuring author contributions to the wikipedia. In *Proceedings of the 4th International Symposium on Wikis*, page 15. ACM, 2008.
- [15] Fabian Flöck and Maribel Acosta. Wikiwho: Precise and efficient attribution of authorship of revisioned content. In *Proceedings of the*

- 23rd international conference on World wide web*, pages 843–854. ACM, 2014.
- [16] Adam Schenker, Abraham Kandel, Horst Bunke, and Mark Last. *Graph-theoretic techniques for web content mining*, volume 62. World Scientific, 2005.
- [17] Andrzej Cichocki and PHAN Anh-Huy. Fast local algorithms for large scale nonnegative matrix and tensor factorizations. *IEICE transactions on fundamentals of electronics, communications and computer sciences*, 92(3):708–721, 2009.
- [18] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.