



UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO

FACULTAD DE POSGRADO

MAESTRÍA EN CIENCIA DE DATOS

**Proyecto de investigación previa la obtención del Grado Académico
de Magister en Ciencia de Datos**

TEMA

Caracterización del perfil de egreso de los profesionales de la
Universidad Técnica Estatal de Quevedo

AUTOR

ING. MARÍA ISABEL CHÉVEZ CASTRO

DIRECTOR

ING. EMILIO RODRIGO ZHUMA MERA MSC.

Quevedo – Ecuador

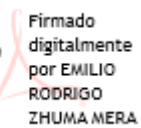
2023-2024

CERTIFICACIÓN

El suscrito certifica que el proyecto para la obtención del grado académico de Magister en Ciencia de Datos, **CARACTERIZACIÓN DEL PERFIL DE EGRESO DE LOS PROFESIONALES DE LA UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO**. De autoría de la Ing. María Isabel Chevez Castro ha sido revisado en todos sus componentes, por lo que autoriza su presentación formal ante el tribunal respectivo.

Quevedo, 2 de abril de 2024.

EMILIO
RODRIGO
ZHUMA
MERA



Firmado digitalmente por EMILIO RODRIGO ZHUMA MERA

Ing. Emilio Zhuma Mera, Ms.C
Director de Tesis

AUTORÍA

Yo, Ing. María Isabel Chevez Castro con CI: 1726622465 declaro que el presente Proyecto de Investigación titulado “**CARACTERIZACIÓN DEL PERFIL DE EGRESO DE LOS PROFESIONALES DE LA UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO**”, es de exclusiva autoría y extendiendo los derechos a la Universidad Técnica Estatal de Quevedo hacer uso del documento como material académico y de consulta.

A handwritten signature in blue ink, appearing to be 'María Isabel Chevez Castro', written in a cursive style.

Ing. Maria Isabel Chevez Castro

CI: 1726622465

DEDICATORIA

A mi madre, mi mejor amiga por ser una gran mujer, valiente y trabajadora, ejemplo de resiliencia, fe y valentía, por su apoyo, soporte y amor incondicional. Quien con su ejemplo me ha enseñado a luchar por mis sueños, sin dejar de lado mis raíces.

A mis hermanas, por ser mi familia, amigas y compañeras de vida, en quienes puedo confiar, llorar y gozar con todas nuestras ocurrencias, a quienes dedico mis logros como su ejemplo y por las que pido a Dios para que sean mucho más bendecidas que yo.

María Isabel Chevez Castro

AGRADECIMIENTO

A la Universidad Técnica Estatal de Quevedo, por abrirme sus puertas y darme todas las facilidades y apoyo para poder estudiar y culminar la maestría. A los docentes de la Facultad de Posgrado, quienes me han compartido sus conocimientos y experiencias, a quienes considero mis colegas y amigos, a los que respeto y deseo lo mejor de la vida.

Al tutor de mi tesis, el ingeniero Emilio Zhuma, por ser mi guía y consejero en la elaboración de esta, con quien he compartido ideas y risas. Quien ha sido una luz en medio de las dificultades y frustraciones que surgieron en el camino. Y quien se merece un reconocimiento por toda su paciencia y apoyo a sus estudiantes.

A todos los grandes profesionales con los que compartí en el programa de maestría, por su amistad, paciencia y risas. Éxitos en su vida profesional.

María Isabel Chevez Castro

PRÓLOGO

En la actualidad la analítica de datos e inteligencia artificial se han convertido en herramientas esenciales para la toma de decisiones no solo para empresas sino también para entes gubernamentales. Es por esta razón, que tanto la minería de datos como la minería de texto toman la batuta en el análisis de grandes volúmenes de datos estructurados y no estructurados (Ruiz Díaz de Salvioni & Armoa, 2023).

El presente trabajo tiene como objetivo analizar una problemática actual: la descripción del perfil de egreso de estudiantes de dos carreras de una universidad ecuatoriana. Este análisis se realizó mediante la aplicación de métodos de minería de texto en los planes analíticos y mallas curriculares, como también en la descripción del perfil profesional demandado por las empresas en un portal de empleo.

El propósito es identificar brechas existentes entre ambos perfiles, que proporcionará información sobre el nivel de empleabilidad de los graduados y servirá como herramienta para la toma de decisiones en la actualización de las mallas curriculares, con el fin de satisfacer las necesidades del mercado laboral. Para lograrlo, este trabajo se encuentra dividido en cinco capítulos, abordando la contextualización de la problemática, una revisión teórica y conceptual, la metodología de investigación, los resultados y discusiones, así como las conclusiones y recomendaciones.



Firmado electrónicamente por:
**ANGEL IVAN TORRES
QUIJIJE**

Ing. Ángel Iván Torres Quijije, MSc.

RESUMEN

En el presente proyecto se analizó la brecha que existe entre el perfil de egreso de los estudiantes de las carreras de Software y Telemática de la Universidad Técnica Estatal de Quevedo y el perfil profesional que demandan las empresas en Hiring Room. Para ello se aplicó técnicas de estadísticas descriptivas, minería de texto y modalización de tópicos con Localización Latente de Dirichlet para encontrar tópicos y palabras claves dentro de textos. Una vez descritos ambos perfiles, se requiere analizar la brecha en cuanto a la capacidad de los estudiantes graduados para ocupar estas vacantes demandadas con ayuda de minería de texto.

Para el análisis del perfil de egreso de las carreras de Software y Telemática se aplicó el método de Localización Latente de Dirichlet o en sus siglas en inglés (LDA) a las mallas curriculares, perfil de egreso y competencias específicas con las que se gradúan los estudiantes. De esta manera se pudo conocer en qué medida los estudiantes se encuentran capacitados para cubrir una vacante en el mercado laboral.

Para el análisis del perfil profesional demandando en Hiring Room, la obtención de los datos se lo realizó mediante la técnica *Web Scraping* o raspado de la Web a los anuncios en Hiring Room, la muestra a considerar fue de 356 anuncios de empleo del presente año en Ecuador, una vez obtenido los datos, se aplicó técnicas de minería de texto y preprocesamiento de texto que faciliten la aplicación de Localización Latente de Dirichlet, describiendo finalmente el perfil profesional.

Palabras claves: Minería de texto, Modelización de Tópicos, Localización Latente de Dirichlet, mercado laboral, software, telemática.

ABSTRACT

This project analyzed the gap that exists between the graduation profile of the students of the Software and Telematics careers of the State Technical University of Quevedo and the professional profile demanded by the companies in the Hiring Room. For this purpose, descriptive statistics techniques, text mining and topic modalization with Dirichlet Latent Localization were applied to find topics and keywords within texts. Once both profiles are described, it is required to analyze the gap in terms of the ability of graduate students to fill these demanded vacancies with the help of text mining.

For the analysis of the graduation profile of the Software and Telematics careers, the Latent Dirichlet Localization (LDA) method was applied to the curricula, graduation profile and specific competencies with which students graduate. In this way it was possible to know to what extent students are qualified to fill a vacancy in the labor market.

For the analysis of the professional profile demanded in Hiring Room, the data was obtained through the Web Scraping technique or scraping of the Web to the ads in Hiring Room, the sample to consider was 356 job ads this year in Ecuador, once obtained the data, text mining techniques were applied and text preprocessing to facilitate the application of Latent Location Dirichlet, finally describing the professional profile.

Keywords: text mining, topic modeling, Latent Dirichlet Allocation, labor market, software, telematics.

TABLA DE CONTENIDO

Contenido

CERTIFICACIÓN	I
AUTORIA.....	II
DEDICATORIA.....	III
AGRADECIMIENTO	IV
PRÓLOGO.....	V
RESUMEN.....	VI
ABSTRACT	VII
TABLA DE CONTENIDO	8
INTRODUCCIÓN	VIII
CAPÍTULO I.....	10
MARCO CONTEXTUAL DE LA INVESTIGACIÓN.....	10
1.1. UBICACIÓN Y CONTEXTUALIZACIÓN DE LA PROBLEMÁTICA.....	1
1.2. SITUACIÓN ACTUAL DE LA PROBLEMÁTICA.....	2
1.3. PROBLEMA DE INVESTIGACIÓN	3
1.3.1. Problemas derivados.....	3
1.4. DELIMITACIÓN DEL PROBLEMA.....	4
1.5. OBJETIVOS	4
1.5.1. Objetivo general	4
1.5.2. Objetivos específicos.....	4
1.6. JUSTIFICACIÓN	6
CAPÍTULO II.....	7
MARCO TEÓRICO.....	7
2.1. FUNDAMENTACIÓN CONCEPTUAL	8
2.1.1. Procesamiento de Lenguaje Natural	8
2.1.2. Minería de texto	8

2.1.3. Modelamiento de tópicos.....	9
2.1.4. Localización latente de Dirichlet.....	9
2.1.5. Estadística Bayesiana	10
2.1.6. Teorema de Bayes	10
2.1.7. <i>Web Scraping</i> – Raspado de Web	10
2.1.8. Mercado laboral	11
2.1.9. Oferta laboral	11
2.1.10. Demanda Laboral	12
2.1.11. Portales de empleo.....	12
2.1.12. Hiring Room	12
2.1.13. Brecha Laboral.....	13
2.1.14. Malla curricular.....	13
2.1.15. Competencias genéricas.....	13
2.1.16. Competencias específicas	13
2.2. FUNDAMENTACIÓN TEÓRICA.....	14
2.2.1. Procesamiento de texto y lenguaje natural.....	14
2.2.2. Minería de Texto	16
2.2.3. Modelización de Tópicos.....	19
2.2.4. Localización Latente de Dirichlet.	21
2.2.5. Aplicaciones del Modelo de Localización Latente de Dirichlet.....	22
2.3. FUNDAMENTACIÓN LEGAL	28
2.3.1. Aspectos Legales del <i>Web Scraping</i> en el Ecuador.....	28
2.3.2. Ley de Comercio Electrónico, Firmas y Mensajes de Datos – 2002:.....	28
2.3.3. Ley Orgánica de Telecomunicaciones - 2015	28
2.3.4. Ley Orgánica de Protección de Datos Personales (LOPDP) – 2021	29
CAPÍTULO III	30
METODOLOGÍA DE LA INVESTIGACIÓN	30
3.2. MÉTODOS UTILIZADOS EN LA INVESTIGACIÓN.....	29

3.2.1. Método Analítico:	29
3.2.2. Método deductivo:.....	29
3.2.3. Método Estadístico:.....	29
3.3. CONSTRUCCIÓN METODOLÓGICA DEL OBJETO DE INVESTIGACIÓN..	30
3.3.1. Población y muestra	30
3.3.2 Técnicas de investigación	32
3.3.3 Instrumentos de Investigación	33
3.4. ELABORACIÓN DEL MARCO TEÓRICO	33
3.5 RECOLECCIÓN DE LA INFORMACIÓN.....	34
1. Portal de empleo Hiring Room	34
2. Mallas curriculares y Planes Analíticos.....	35
3.6 PROCESAMIENTO Y ANÁLISIS	36
3.6.1. Perfil demandando por las empresas en el portal de empleo Hiring Room	36
CAPÍTULO IV	38
RESULTADOS Y DISCUSIÓN	38
4.1. APLICACIÓN DE LAS TÉCNICAS <i>WEB SCRAPING</i> Y MODELIZACIÓN DE TÓPICOS A LOS ANUNCIOS DE EMPLEO DEL PORTAL HIRING ROOM, PARA LA DESCRIPCIÓN DE LOS PERFILES DE VACANTES DEMANDADAS POR LAS EMPRESAS PÚBLICAS Y PRIVADAS	38
4.1.1. Aplicación de <i>Web scraping</i> a los anuncios de empleo del portal Hiring Room	38
4.1.2. Aplicación de la técnica Localización Latente de Dirichlet (LDA) para descripción del puesto, requisitos de educación, conocimiento, experiencia y competencias generales.	46
4.2. ANÁLISIS DEL PERFIL DE EGRESO DE LAS CARRERAS DE SOFTWARE Y TELEMÁTICA CON BASE EN LAS COMPETENCIAS ESPECÍFICAS MEDIANTE LA MODELIZACIÓN DE TÓPICOS.	51
4.2.1. Análisis descriptivo de la Malla Curricular de la carrera de Software	51
4.2.2. Análisis de la Malla curricular de la carrera de Telemática	53

4.2.3. Aplicación LDA en la descripción de las competencias específicas de ambas carreras.....	53
4.3. ANÁLISIS DE LA BRECHA EXISTENTE ENTRE AMBOS PERFILES HALLADOS, MEDIANTE HERRAMIENTAS DE MINERÍA DE TEXTO.	56
CAPÍTULO V	63
CONCLUSIONES Y RECOMENDACIONES	63
5.1. Conclusiones.....	61
5.2. Recomendaciones.....	62
REFERENCIAS BIBLIOGRÁFICAS.....	63

Índice de tablas

Tabla 1: Ciudades en las que se ubican las empresas que publican anuncios de empleo	40
Tabla 2: Modalidad de trabajo de acuerdo con la subárea de trabajo	44
Tabla 3: Perfiles hallados por Localización Latente de Dirichlet para las vacantes demandadas en el portal de empleo Hiring Room	49
Tabla 4: Cantidad de créditos por semestre para la carrera de Software	52
Tabla 5: Estadísticos descriptivos para horas y créditos de la carrera de Software.....	52
Tabla 6: Estadísticos descriptivos de número de horas y créditos de la carrera de Telemática.....	53
Tabla 7: <i>Perfiles hallados por medio de Localización Latente de Dirichlet para el perfil de egreso de los estudiantes de ambas carreras</i>	55

Índice de ilustraciones

Ilustración 1: Conceptualización de los pasos de la minería de texto	17
Ilustración 2: Conceptualización del Modelo Localización Latente de Dirichlet	21
Ilustración 3: Esquematación del Modelo Localización Latente de Dirichlet(LDA)..	22

Ilustración 4: Tópicos creados por LDA para el estudio de caso en el contexto educativo	27
Ilustración 5: Proceso de la extracción de los anuncios de empleo mediante Web Scraping	38
Ilustración 6: <i>Top 15 de las empresas que más anuncios de empleo publican en Hiring Room en la búsqueda de profesionales del área de las TIC's</i>	39
Ilustración 7: Ciudades a realizar las actividades de acuerdo con los anuncios de empleo	41
Ilustración 8: Anuncios de empleo de acuerdo con las subáreas de Tecnología, Sistemas y Telecomunicaciones	42
Ilustración 9: Tiempo de la Jornada Laboral	43
Ilustración 10: Modalidad de trabajo.....	43
Ilustración 11: Años de experiencia laboral requeridos	45
Ilustración 12: Procedimiento lógico de la modelización de tópicos por Localización Latente de Dirichlet	47
Ilustración 13: Procedimiento del algoritmo de Localización Latente de Dirichlet.....	48
Ilustración 14: <i>Extracción de la información de los planes analíticos mediante Web Scraping empleando librerías de R-studio y almacenándolos en una base de datos de Excel.</i>	54
Ilustración 15: Proceso de la librería wordcloud junto con sus funciones de minería de texto	57
Ilustración 16: Gráfica de nube de palabras para analizar las coincidencias más frecuentes entre el perfil demandado en las vacantes de Hiring room y el perfil de egreso de los graduados de las carreras de software y telemática.	58
Ilustración 17: <i>Nube de palabras para hallar la brecha existente entre el perfil de egreso y las vacantes demandadas.</i>	60

Índice de Anexos

Anexo 1: Informe Compilatio – Antiplagio	1
Anexo 2: Malla curricular de Software	73
Anexo 3: Malla Curricular de la Carrera de Telemática	74
Anexo 4: Ejemplo de la obtención de las probabilidades Beta y gammas de que una palabra clave pertenezca a alguno de los tópicos establecidos en los objetivos de cada asignatura y las probabilidades de que cada documento pertenezca a un tópico	75
Anexo 5: Creación de tópicos con el Modelo de Localización Latente de Dirichlet a las descripciones de las vacantes	76
Anexo 6: Creación de tópicos por medio de Localización Latente de Dirichlet para los requisitos de educación	78
Anexo 7: Creación de tópicos mediante Localización Latente de Dirichlet a los requisitos de conocimiento	81
Anexo 8: Creación de tópicos mediante Localización Latente de Dirichlet para habilidades blandas	84
Anexo 9: Creación de tópicos mediante LDA a las descripciones de los planes analíticos	85
Anexo 10: Creación de tópicos mediante LDA a los objetivos generales de los planes analíticos de ambas carreras	87
Anexo 11: Creación de tópicos mediante LDA para los objetivos específicos de los planes analíticos de cada carrera	89
Anexo 12: Creación de tópicos mediante LDA a los resultados de los planes analíticos de ambas carreras	90
Anexo 13: Creación de tópicos mediante LDA a los temas y subtemas de los planes analíticos de ambas carreras	91
Anexo 14: Ejemplo de un anuncio de empleo del portal Hiring Room	92
Anexo 15: Ejemplo de plan analítico de la carrera de Software	93

INTRODUCCIÓN

La inserción de un recién egresado de la universidad en el mercado laboral es cada vez más compleja, debido a múltiples factores como la globalización y el auge del desarrollo tecnológico, haciendo que tanto estudiantes como universidades se encuentren en constante aprendizaje y actualización. Otro factor, es la alta e innovadora especialización de profesionales de recursos humanos de empresas públicas y privadas que con metodologías sofisticadas buscan encontrar a los mejores perfiles para que formen parte de las empresas, perfiles que demandan ciertas habilidades que muchas de las veces no son aprendidas en el salón de clases (Solé-Moro et al., 2018).

La misión fundamental de la universidad es, por lo tanto, proporcionar una formación suficiente en cuanto a conocimientos y habilidades que permitan la empleabilidad y competitividad de sus egresados en el mercado laboral. Es por ello, que deben actualizar constantemente sus mallas curriculares y los resultados del aprendizaje de cada cátedra de las carreras profesionales (Echeverría-Villafuerte y Villarruel-Meythaler, 2022).

Todo lo anterior, ligado al contexto socioeconómico que vive el Ecuador en la última década, requiere analizar de manera sistemática y rigurosa acerca de la brecha que existe entre la calidad de profesionales que gradúa la Universidad Técnica Estatal de Quevedo versus los perfiles profesionales demandados por las empresas públicas y privadas.

El fin de esta investigación es, por tanto, comparar el perfil profesional de las carreras de Software y Telemática de la Universidad Técnica Estatal de Quevedo con las vacantes

demandadas en un portal de empleo para la determinación de la brecha existente entre ambos.

Se ha estructurado la investigación de la siguiente manera: En primer lugar, se plantea una revisión bibliográfica acerca de la problemática a estudiar y de la minería y procesamiento de texto mediante la Modelización de Tópicos por Localización Latente de Dirichlet y la aplicación de la técnica *Web Scraping* para la obtención de datos. En segundo lugar, se realizó un análisis descriptivo de las fuentes de información (Hiring Room, planes analíticos y perfil de egreso de las carreras de Software y Telemática de la Universidad Técnica Estatal de Quevedo). En tercer lugar, se aplicó minería de texto y modelización de tópicos. Continuando con los resultados en cuanto a la brecha entre el perfil profesional demandado por las empresas en la página Hiring Room y el perfil de egreso de las carreras de software y telemática de la Universidad Técnica Estatal de Quevedo. Por último, se presentan las conclusiones alcanzadas con la investigación y recomendaciones para trabajos futuros.

CAPÍTULO I

MARCO CONTEXTUAL DE LA INVESTIGACIÓN

1.1. UBICACIÓN Y CONTEXTUALIZACIÓN DE LA PROBLEMÁTICA

La Universidad cumple un rol esencial en la sociedad, ya que se encarga de preparar a los profesionales adecuados para satisfacer las necesidades del mercado laboral. Los principales objetivos de la Universidad son brindar una formación y cualificación suficiente que garantice la empleabilidad y la competitividad de sus titulados, permitiéndoles alcanzar una inserción laboral exitosa (Varea, 2021).

Para (Wang, 2022) la malla curricular sirve como base para alcanzar los objetivos educativos, por lo que la renovación de estos junto con planes de estudios es un tema debatido entre académicos y especialistas en la educación. Pues muchos piensan que a mayor carga en el plan de estudios es más útil en el mercado laboral, sin embargo, no siempre se cumplen esta hipótesis (Seah et al., 2020).

Por ejemplo, en un análisis de aptitudes y cualidades profesionales realizado por (Ridge, 2024), se llegó a la conclusión de que las empresas mayormente exigen: experiencia, trabajo bajo presión, capacidad de comunicación, uso de programas y nuevas tecnologías, trabajo en equipo, creatividad y análisis de datos. De las cuales muchas de estas no son impartidas en clases, y que muy difícilmente la carga adicional en la malla curricular proporcione habilidades blandas. En vista de esto, algunas universidades ofrecen cada vez más cursos prácticos e interdisciplinarios. Sin embargo, sigue existiendo un desfase entre los planes de estudios y los perfiles demandados en el mercado (Wang, 2022).

Es crucial dar un seguimiento a los estudiantes graduados y su inserción laboral para que las universidades puedan conocer las demandas emergentes. En base a esto, la enseñanza superior debería incentivar a que los estudiantes desarrollen técnicas y habilidades como

la reflexión crítica, la resolución de problemas y el autoaprendizaje, ayudando a prepararse no solo para las tareas actuales, sino también para anticipar y sugerir nuevas ideas (Sierra, 2022).

El papel de la universidad en la formación de profesionales en toda Latinoamérica se ha visto afectada debido a diversos problemas culturales y sociopolíticos, uno de ellos es la fuga de conocimiento, pues por la falta de oportunidades laborales, las personas suelen migrar hacia otros países desarrollados o con una mejor economía (Solé-Moro et al., 2018).

1.2. SITUACIÓN ACTUAL DE LA PROBLEMÁTICA

En el análisis de la situación actual de la problemática se encuentran dos entes, por un lado, el papel de la universidad y por otro el mercado laboral ecuatoriano, a continuación, la descripción de ambos.

La Universidad Técnica Estatal de Quevedo, es una universidad pública ubicada en el cantón Quevedo de la provincia de Los Ríos, misma que en sus 39 años de existencia ha graduado a muchos profesionales y actualmente es pionera en la formación de profesionales de calidad.

Una de las facultades de la Universidad Técnica Estatal de Quevedo es la Facultad de Ciencias de la Ingeniería, compuesta por 8 carreras técnicas y tecnológicas en las que se encuentran las carreras de Ingeniería en Software y Telemática, objeto de análisis en la presente investigación. Cada una de las carreras cuenta con mallas curriculares bien estructuradas, planes analíticos, sílabos, etc. Tienen una duración de 10 semestres, y además propone y comparte cursos complementarios a sus estudiantes, con el fin de adecuarlos a las demandas del mercado laboral. Es importante señalar que en el tema de

la investigación no se especifican estas carreras por motivos burocráticos, y requerimientos internos de la institución.

La realidad del mercado laboral ecuatoriano es que existe una tendencia en las últimas décadas de haber un número elevado de graduados en todas las ramas, pero pocas oportunidades laborales adecuadas a sus habilidades (Bernal-Yamuca et al., 2022), por lo que en consecuencia los profesionales se han visto en la necesidad de realizar cursos, diplomados, maestrías, doctorados que les permitan competir por esos limitados puestos de trabajo.

En la actualidad, Ecuador vive en una recuperación del empleo, no obstante, el mercado laboral está mayormente constituido por profesionales o estudiantes jóvenes con empleo precario, subempleo, o trabajando en ramas por lo que no fueron preparados (Bernal-Yamuca et al., 2022). Por lo que el papel de la universidad en el Ecuador tiene un trabajo duro en cuanto a la formación de profesionales que compitan en este mercado.

1.3. PROBLEMA DE INVESTIGACIÓN

A todo lo anteriormente explicado, surge la siguiente problemática:

¿Cómo el perfil de egreso de los profesionales de las carreras de Software y Telemática de la Universidad Técnica Estatal de Quevedo cubre con las exigencias de las vacantes demandadas en el portal de empleo Hiring Room?

1.3.1. Problemas derivados

- ¿Cómo determinar las características que exigen las empresas públicas y privadas en el portal de empleo Hiring Room?
- ¿Cómo identificar las competencias específicas que tienen los profesionales de las carreras de Software y Telemática de la Universidad Técnica Estatal de Quevedo?

- ¿De qué manera se puede analizar la brecha que existe entre los profesionales de las carreras de Software y Telemática de la Universidad Técnica Estatal de Quevedo y las vacantes demandadas en el portal de empleo Hiring Room?

1.4. DELIMITACIÓN DEL PROBLEMA

- CAMPO: Ciencias de la ingeniería y tecnología, TIC's
- ÁREA: Tecnologías de la Información y Comunicaciones
- LÍNEA: Ciencia de Datos, Educación
- LUGAR: Ecuador
- TIEMPO: Septiembre 2023- Marzo 2024

1.5. OBJETIVOS

1.5.1. Objetivo general.

Comparar el perfil profesional de las carreras de Software y Telemática de la Universidad Técnica Estatal de Quevedo con las vacantes demandadas en Hiring Room basándose en modelización de tópicos para la determinación de la brecha existente entre ambos.

1.5.2. Objetivos específicos

- Aplicar técnicas de *Web Scraping* y modelización de tópicos a los anuncios de empleos del portal Hiring Room, para la descripción de perfiles de vacantes demandadas por empresas públicas y privadas.
- Analizar el perfil de egreso de las carreras de Software y Telemática con base en las competencias específicas mediante la modelización de tópicos.

- Analizar la brecha existente entre los tópicos identificados con el modelo de Localización Latente de Dirichlet en las competencias específicas y los anuncios de empleo, mediante técnicas de minería de texto que brinde una referencia sobre la funcionalidad de las mallas curriculares y grado de empleabilidad de los graduados de las carreras analizadas.

1.6. JUSTIFICACIÓN

El análisis de la inserción laboral de los estudiantes recién graduados de una universidad es importante tanto para la calificación de la calidad de educación de la Institución de Educación Superior como la competitividad exitosa de los profesionales en el mercado laboral. Este análisis ayuda a la planificación de mallas curriculares y planes de estudio que cubran con las necesidades demandadas por el mercado (Castillo y Aguilar, 2020).

El cálculo de la brecha que existe entre la oferta y la demanda laboral es una medida importante para la universidad, pues si esta brecha tiende a ser alta, significa que la calidad de educación de un Instituto de Educación Superior no es buena, o su planificación es inválida. Mientras que, si la brecha es baja, significa que la educación es de calidad media alta, y que se debe considerar otros aspectos que exige el mercado laboral (Paterson, 2023).

La aplicación de las técnicas de Aprendizaje Automático, Ciencia de Datos, Minería de Datos, pueden llegar a ser herramientas fundamentales para la comprensión de tópicos relacionados a la competitividad de la universidad y los resultados obtenidos por ella. Como, por ejemplo, si los profesionales graduados por ella logran competir y cubrir necesidades de las empresas.

La presente investigación pretende ser ayuda para la Universidad Técnica Estatal de Quevedo para poder conocer en qué medida sus planes curriculares gradúan profesionales que puedan competir en el mercado con éxito. Además, busca brindar una guía para la aplicabilidad de la Ciencia de datos en su toma de decisiones. Y ser fuente para posibles investigaciones futuras.

CAPÍTULO II

MARCO TEÓRICO

2.1. FUNDAMENTACIÓN CONCEPTUAL

A continuación, se enlistan conceptos relacionados con el problema a investigar, con el fin de aclarar las temáticas a utilizar.

2.1.1. Procesamiento de Lenguaje Natural

Es una rama de la inteligencia artificial cuyo objetivo es investigar, facilitar, desarrollar métodos que faciliten la comunicación entre humanos y máquinas utilizando el lenguaje de las personas en cualquier idioma (Diaz-Rubiano, 2022), con datos no estructurados como el habla y el texto de la comunicación del ser humano.

2.1.2. Minería de texto

Es un enfoque común del Procesamiento de Lenguaje Natural, la cual implica un tratamiento de datos textuales. Su objetivo principal es desvelar patrones y tendencias ocultas en fuentes de datos textuales como patentes, redes sociales o cualquier tipo de documento, brindando a las empresas información útil para la toma de decisiones posicionándolas en el mercado (Lupi y Mabkhot, 2023).

La diferencia entre la minería de datos y la minería de texto radica en que, la primera trata cualquier tipo de información, desde millones de registros en una base de datos al análisis de imágenes (Prabha y Sardana, 2023), mientras que la segunda, es basada en la lingüística computacional y el procesamiento de texto para obtener nuevo conocimiento (Echeverría-Villafuerte y Villarruel-Meythaler, 2022).

2.1.3. Modelamiento de tópicos

Es una técnica usada en el Procesamiento de Lenguaje Natural y Minería de Texto, cuyo objetivo se basa en descubrir patrones ocultos y agruparlos en categorías o tópicos coherentes (Annisa y Surjandari, 2019). Esta técnica considera los temas como mezcla de temas probabilísticos, donde el tema o tópico se representa por una distribución de probabilidad sobre las palabras específicas (Echeverría-Villafuerte y Villarruel-Meythaler, 2022).

(Blei et al., 2003) define a la modelización de tópicos como una serie de técnicas de procesamiento de Lenguaje Natural que emplea técnicas informáticas y estadísticas para comprimir y consolidar el contenido de grandes cantidades de texto. El cual, mediante las aplicaciones de técnicas estadísticas en textos originales pueden describir una estructura semántica subyacente no observada (Luque et al., 2021), es decir, este modelo asume que en cualquier colección de texto hay un conjunto de tópicos o temas ocultos que organizan el conjunto de documentos.

2.1.4. Localización latente de Dirichlet

Técnica de modelado de texto más utilizado la cual usa una distribución multinomial de palabras en temas. Este algoritmo tiene sus bases probabilísticas en la Estadística Bayesiana e Inferencia Estadística Bayesiana (Islam et al., 2022).

Este algoritmo funciona bajo la lógica que cada documento tiene una mezcla de varios temas que son asignados a través del mismo algoritmo, suponiendo que la distribución de cada tema proviene de una distribución de Dirichlet. Esto significa que el algoritmo permite que cada documento pertenezca a diferentes tópicos al mismo tiempo, con un peso diferente para cada uno de ellos. En resumen, cada documento o bolsa de palabras

tiene más probabilidad de pertenecer a un determinado tópico que a otro (AnalyStats, 2019).

2.1.5. Estadística Bayesiana

A diferencia de la estadística frecuentista (la más antigua y usada), la Estadística Bayesiana incorpora información externa o exógena a un evento actual, es decir, la probabilidad de un evento actual se ve afectado por información de sucesos pasados.

A partir de esta surge el Teorema de Bayes, la cual describe la probabilidad de ocurrencia de un determinado evento se ve condicionado por sucesos pasado (Diaz-Rubiano, 2022).

2.1.6. Teorema de Bayes

El teorema de Bayes es uno de los términos más usados en la Estadística Bayesiana, este teorema describe que la probabilidad de ocurrencia de un evento se encuentra condicionado por una serie de eventos pasados (Terreros-Ríos et al., 2019), este viene dado por la siguiente expresión:

Ecuación 1: Teorema de Bayes

$$P(Y) = P(Y|X_1)P(X_1) + P(Y|X_2)P(X_2) + \dots + P(Y|X_n)P(X_n)$$

$$P(Y) = \sum P(Y|X_i)P(X_i)$$

Elaboración: Autora

La probabilidad del evento Y es igual a la probabilidad de Y dado X_i , donde, Y es el evento a predecir y X_i son los eventos pasados (Diaz-Rubiano, 2022).

2.1.7. Web Scraping – Raspado de Web

Técnica de la Ciencia de Datos que permite obtener aquella información útil para un proyecto de datos que se está disponible en internet (Laia & Calvo, 2019), de manera automatizada optimizando recursos como el tiempo y minimizando errores. Así, el raspado de Web consiste en “la construcción de un agente que permita descargar, analizar y organizar datos procedentes de internet, de forma automática” (Laia & Calvo, 2019).

Es importante tener algunas consideraciones legales y éticas al hablar del *Web Scraping*, las personas que emplean esta técnica deben tener en cuenta las leyes de seguridad cibernética actuales aplicadas a sus actividades. Por lo tanto, el responsable debe cumplir con los términos de uso establecidos por el sitio web del que se desea extraer los datos (Stenhouse, 2018). Entre los softwares más utilizados para aplicar esta técnica (Dogucu & y Çetinkaya-Rundel, 2020) y (AVI Networks, s. f.) siguieren Python como primera opción seguido de R, además, indican que *Web scraping* debería ser incorporada dentro de los planes de estudio de una manera efectiva y eficiente, en especial en niveles de programas académicos de estadística y ciencia de datos.

2.1.8. Mercado laboral

El mercado laboral está conformado por un conjunto de factores tales como la Población Económicamente Activa, niveles de instrucción, la demanda y dinámica de los sectores productivos y sociales, ingresos y otros más, es decir, se encuentra conformado por la oferta y le demanda de los sectores productivos de un país específico (Bernal-Yamuca et al., 2022).

2.1.9. Oferta laboral

Representa la participación de los trabajadores en el mercado laboral. Es la cantidad de horas de trabajo que las familias están dispuestos trabajar (Bernal-Yamuca et al., 2022).

Se encuentra conformado con el nivel de instrucción, conocimiento, experiencia laboral que los ciudadanos cuentan para cubrir con los puestos de trabajo que exigen la empresa a cambio de un ingreso.

2.1.10. Demanda Laboral

La demanda de trabajo se define como la cantidad de trabajadores que las empresas quieren contratar para poder producir bienes y servicios. En la mayoría de los casos, los empresarios estarán dispuestos a contratar a nuevos empleados siempre y cuando los ingresos generados por ellos sean superiores al salario que deben pagar (Bernal-Yamuca et al., 2022). Las empresas además exigen ciertas características mínimas que el candidato ideal debe cumplir.

2.1.11. Portales de empleo

Son plataformas en línea, las cuales funcionan como intermediarios donde las empresas publican ofertas de trabajo y los candidatos pueden postularse a estos siempre y cuando cumplan con las características que las empresas necesitan (Valle Gutiérrez, 2021).

El mismo autor, quien en su estudio indica que entre los portales más conocidos en el Ecuador están: CompuTrabajo, Multitabajos, Porfin empleo, Socio empleo, entre otras.

2.1.12. Hiring Room

Por la gran cantidad de portales de empleo, la empresa Navent creó un software de reclutamiento llamado “Hiring Room”, el cual ofrece algunas ventajas como la reducción de hasta el 50% el tiempo de reclutamiento, además genera métricas sobre el desempeño en el área de Recursos Humano, propone análisis de indicadores y filtros para poder encontrar al perfil indicado (Info Capital Humano, s. f.), cabe señalar también que Hiring Room trabaja de a mano con Multitabajos, uno de los portales de empleo más usados y reconocidos en el Ecuador. Cada vez son más las empresas que optan por Hiring Room.

2.1.13. Brecha Laboral

Se refiere a la diferencia existente entre las capacidades de una profesional obtenida por un ente educativo frente a las exigencias que demandan las empresas para cubrir un puesto de trabajo (Wang, 2022).

2.1.14. Malla curricular

Una malla curricular es un plan de estudios organizado y estructurado que define los cursos o asignaturas, contenido, competencias, créditos académicos, métodos de evaluación y requisitos necesarios para completar un programa educativo específico (Dympna Kelly et al., 2015). Según (Wang, 2022) su estructura básica contiene

1. Introducción y Objetivos del Programa
2. Perfil del Estudiante
3. Estructura del Plan de Estudios
4. Descripción de Cursos o Asignaturas
5. Créditos Académicos
6. Evaluación

2.1.15. Competencias genéricas

Términos relacionados a la formación académica de un estudiante, también conocidas como competencias transversales, habilidades blandas o habilidades socioemocionales. Este tipo de competencia son esenciales para la vida personal y profesional. Por ejemplo, comunicación afectiva, trabajo en equipo, pensamiento crítico y resolución de problemas, creatividad, adaptabilidad, entre otras más (Dympna Kelly et al., 2015).

2.1.16. Competencias específicas

Al igual que las competencias genéricas es un término relacionado a la formación académica de un estudiante, también conocidas como competencias técnicas o competencias disciplinares (Dympna Kelly et al., 2015), son aquellas habilidades y conocimientos especializados relacionados a un campo de estudio particular. Este tipo de competencia es esencial para el desempeño efectivo de una carrera descritas por una malla curricular (Ilyés y Seb Anna, 2023).

2.2. FUNDAMENTACIÓN TEÓRICA

En el auge de la Ciencia de Datos y el Big Data, surgen términos como el procesamiento de lenguaje natural, minería de texto y modelización de tópicos, los cuales han sido herramientas para el análisis de problemáticas y fenómenos sociales, económicos y tecnológicos. La modelización de tópicos, en especial la aplicación del algoritmo de Localización Latente de Dirichlet ha sido uno de los algoritmos más usados en lo que respecta a procesamiento del lenguaje natural y análisis de sentimientos.

En el análisis de la contextualización teórica se realizó una síntesis de todas las investigaciones referentes a la Modelización de tópicos usando la Localización Latente de Dirichlet aplicadas a diferentes escenarios como el educativo, salud, empresarial.

2.2.1. Procesamiento de texto y lenguaje natural

(Valle Gutiérrez, 2021) en su investigación, indica que el lenguaje natural surge por la necesidad de almacenar y analizar los datos de forma no convencional, como se ha venido haciendo en el pasado. Debido al auge del Big Data y a la potencia computacional actual el procesamiento de texto y análisis de sentimientos son necesarios para comprender los millones de datos que se producen a diario, por medio de las redes sociales, transacciones, registros e internet. Por otra parte, (Castillo y Aguilar, 2020) hacen referencia a que el procesamiento de lenguaje natural es la base para la creación de técnicas y algoritmos

para el análisis de texto, como la minería de texto, modelización de tópicos y otras técnicas más, las cuales combinadas con las estadísticas llevan a la comprensión del comportamiento de los datos (Maier et al., 2018). Es por ello que, estos escritores en sus publicaciones brindan una gama de técnicas de procesamiento de texto aplicadas a diferentes contextos, entre las que se utilizó están la Modelización de tópicos y el modelo de Localización Latente de Dirichlet.

Además, diversos autores señalan que el procesamiento de texto se lo debe realizar bajo los siguientes pasos (Echeverría-Villafuerte y Villarruel-Meythaler, 2022) y (Valle Gutiérrez, 2021), los cuales permiten a los algoritmos determinar tópicos o relaciones entre palabras que a simple vista y por intuición del lector son difíciles encontrar. Primero, se encuentra la Tokenización: Que es el paso en el que al texto general se lo subdivide en pequeños fragmentos llamados “Tokens”, en la mayoría de los casos estos pueden ser palabras o grupos de palabras. Estos fragmentos ayudan a los modelos a interpretar los tópicos dentro de un grupo de palabras. Segundo, Eliminación de stopwords, paso en el cual se trata de eliminar aquellas palabras que aparecen en casi todos los documentos y no son consideradas ya que dificultan el proceso de recuperación de la información, entre estas palabras están: Artículos, pronombres, preposiciones, conjunciones. Tercero, Lematización o stemming, en este paso las palabras se reducen a su forma raíz o base eliminando afijos derivativos de los extremos de la palabra. Este paso facilita al procesamiento de lenguaje natural reconociendo grupos de palabras relacionados a su raíz. Por ejemplo:

- Palabra Original: Corriendo

Stem: Corre

- Palabra Original: Hablaremos

Stem: Habl

Tercero, el etiquetado de partes de un discurso (*Parto f Speech* o PoS) hace referencia a la clasificación de las palabras por categoría gramatical, como sustantivo, pronombre, etc. Para determinar cómo las palabras se combinan con otras palabras para formar una estructura gramatical coherente. Cuarto, *Parsing* o Análisis sintáctico, en este último paso se analiza una parte de un texto para determinar su estructura lógica, mediante un parser o analizador de lenguaje natural, se encuentra formado por: un preprocesador, analizador morfológico y un diccionario. “Se encuentra representado por un árbol, cuyas hojas y nodos representan las palabras del texto y las estructuras sintácticas respectivamente (Echeverría-Villafuerte & Villarruel-Meythaler, 2022).

De acuerdo con los autores anteriormente citados, el procesamiento de lenguaje natural se encuentra resumido en los 4 pasos anteriores, los cuales permiten a las máquinas comprender el lenguaje humano y su comportamiento sin importar el idioma o su estructura.

2.2.2. Minería de Texto

Para (Sans et al., 2022) la minería de texto surge a “partir del Descubrimiento de Conocimiento a partir de Bases de Datos” (KDD en sus siglas en inglés), mientras que para (Diaz-Rubiano, 2022) la minería de texto se basa de un área, la cual se encarga del estudio de información digital y textual.

Algunos autores proponen ciertos pasos en la cual se debe desarrollar la minería de texto. Primero, determinar el propósito del estudio de la minería de texto aplicada a cualquier realidad. Segundo, recolectar, identificar y validar la información necesaria, para este paso (Laia & Calvo, 2019) propone algunas formas de recolección de datos para la aplicación de técnicas de minería de texto como lo es el *Web Scraping*, el cual ha ayudado

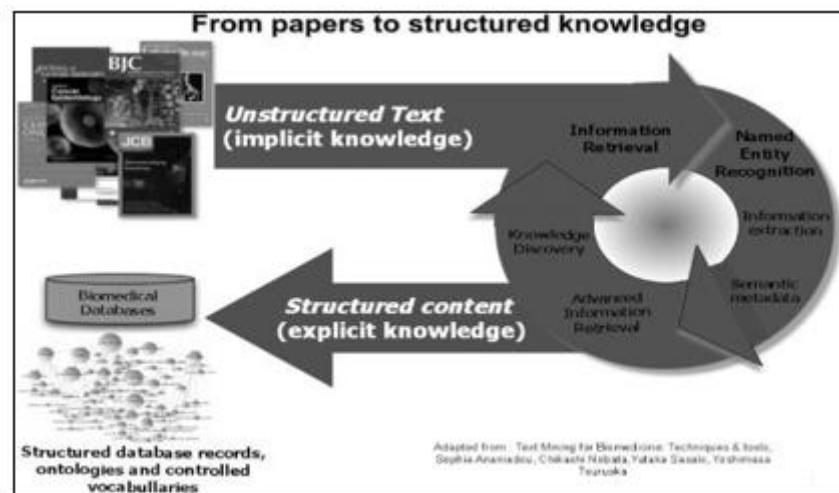
a la obtención de datos como en redes sociales y portales de empleo, así lo han aplicado los siguientes autores (Vernazza, 2020), (Valle Gutiérrez, 2021), (Castillo & Aguilar, 2020), (Islam et al., 2022), (Annisa & Surjandari, 2019).

Tercero, realizar el preprocesamiento de texto, en donde se eliminan palabras no útiles, se buscan las raíces de las palabras, etc., en otras palabras, se puede aplicar los pasos que se realizan en el proceso de lenguaje natural o procesamiento de texto, anteriormente detallados.

Cuarto, extracción y análisis de clases, relaciones, o asociaciones, en esta etapa los documentos se representan como espacios vectoriales, además se encuentran representados por alguna estructura informática – estadística que faciliten su análisis.

Quinto, representación de los resultados, a través de resúmenes, visualizaciones que faciliten su interpretación. A continuación, una ilustración que facilita el entendimiento de los pasos.

Ilustración 1: Conceptualización de los pasos de la minería de texto



Fuente: (Contreras - Barrera, 2014)

La figura muestra el proceso de la minería de texto, partiendo desde un conocimiento implícito en el texto a un conocimiento explícito con un contenido estructurado.

(Buenano-Fernandez et al., 2020) indica que a pesar de que la Ciencia de Datos ha creado algoritmos que ayuden al análisis de los datos por medio de la minería de datos y aprendizaje automático supervisado y no supervisado, existe aún un desfase en lo que respecta al análisis de texto que millones de fuentes como redes sociales y la misma internet producen a diario, (Diaz-Rubiano, 2022) resume que las redes sociales cada minuto producen millones de millones de datos alrededor del mundo, datos que no han sido estudiados a profundidad. Partiendo de esta premisa, parte la minería de texto y sus modelos como herramientas útiles para la interpretación de datos textuales, como lo es la modelización de tópicos.

El análisis de texto además de ser un mundo que no ha sido indagado a profundidad, la obtención de sus datos en muchas de las ocasiones se ha convertido en un reto (Diaz-Rubiano, 2022). Por ejemplo, El análisis de preferencias en las principales ciudades colombianas en la red social Twitter, la obtención de los datos se realizó por medio de la API de esta red, sin embargo, la obtención de la data se convierte en un reto cuando no se dispone de alguna API, la cual en muchos de los casos tiene un valor monetario. Así surgen algunas técnicas como *Web Scraping*, como se explicó en unos párrafos anteriores. El objetivo de esta técnica es obtener la data que se encuentra en el internet, como, anuncios de empleo, comentarios acerca de alguna película, música, etc. Sin embargo, como lo indica (Dogucu y Çetinkaya-Rundel, 2020) es importante considerar aspectos legales al momento de optar por esta técnica.

Es así como la obtención de la data para la aplicación de minería de texto a través del *Web Scraping* ha sido utilizado para estudios como: Análisis de sentimientos, análisis

preelectoral, análisis de las publicaciones medicas realizadas en una revista, análisis de perfiles que demandan las empresas en portales de empleo, tendencias a suicidio analizadas en redes sociales, entre otros más.

2.2.3. Modelización de Tópicos

Una de las técnicas de la minería de texto y de procesamiento de lenguaje natural es la Modelización de Tópicos. De acuerdo con (Prabha y Sardana, 2023) la Modelización de Tópicos ha sido explorado en diversos estudios relacionados a la minería de texto, por lo que es una de las técnicas más populares para identificar tópicos principales dentro del cuerpo de un texto. El modelamiento de texto describe cada documento como una colección probabilística de tópicos y cada tópico como una distribución probabilística de diferentes palabras. Este método ha sido ampliamente aplicado en áreas como Ingeniería de Software, Ciencia Biomédica, Literatura y publicaciones científicas (Prabha & Sardana, 2023).

(Terreros-Ríos et al., 2019) explica que la Modelización de Tópicos se da bajo la lógica del Teorema de Bayes, el cual describe que la probabilidad de ocurrencia de un evento se encuentra condicionado por una serie de eventos pasados.

(Buenano-Fernandez et al., 2020) y (Pietsch & Lessmann, 2018) señalan que la Modelización de Tópicos forma parte del aprendizaje automático no supervisado, pues este, implica un proceso de inferencia en el que no se conoce la variable respuesta. Más bien, el modelo en sí es que el que entrega respuestas que no son conocidas por el investigador. Existen varios métodos dentro de lo que respecta Modelización de Tópicos, dentro de los más utilizados y comunes destacan tres métodos con sus respectivos objetivos muy bien diferenciados (Laboratorio de Datos, 2021), los cuales son: Análisis Latente Semántico (LSA), el cual es utilizado para reducciones en la dimensionalidad de un texto

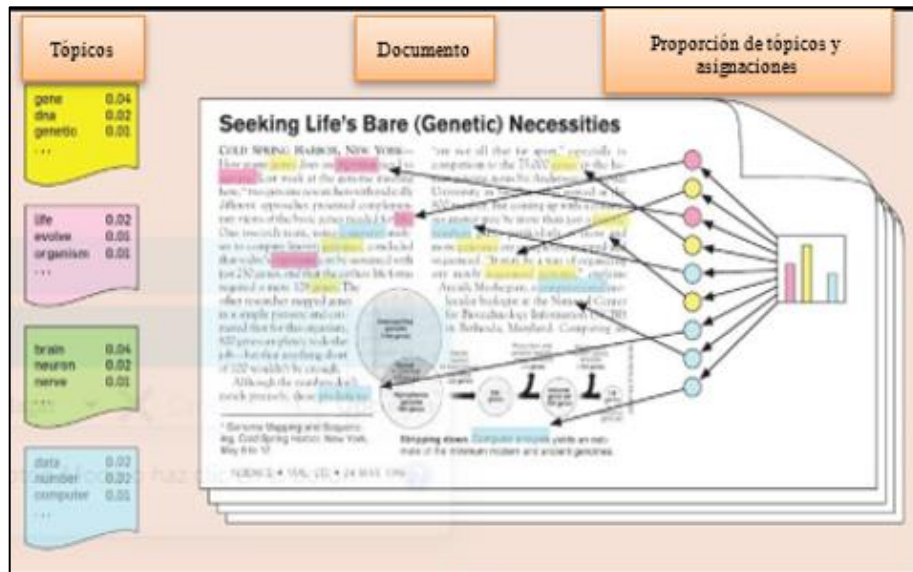
con una pérdida mínima de información, por lo que está relacionado al Análisis de Componentes Principales, sin embargo, este algoritmo no brinda una explicación clara de la creación de tópicos como lo hacen otros.

El segundo método, llamado Factorización Matricial no Negativa (NMF) es una descomposición de datos matriciales, la cual requiere la no-negatividad de cada uno de los componentes, manejando solo datos positivos que combinados da una fuente (tópico) que no necesariamente deben ser ortogonales. Su objetivo se basa en la reducción de la dimensionalidad y la extracción de características. Una de las desventajas de este modelo, es que la cantidad de tópicos D fija un nivel de absorción, es decir, aquellos que el algoritmo suele encontrar tópicos de tamaño similar, es decir absorbe aquellos tópicos pequeños o puede dividir a los tópicos muy grandes. También es importante señalar que los resultados de NMF son difíciles de interpretar en términos de tópicos o categorías.

El tercer método es Localización Latente de Dirichlet (LDA), tiene un enfoque probabilístico, su objetivo se basa en el modelado de temas y clasificación de documentos para descubrir los temas latentes en un conjunto de documentos y asignar documentos a estos temas. Por su enfoque probabilístico, modela la probabilidad de que un documento pertenezca a un tema, y la probabilidad de que una palabra este asociado a un tema.

En vista a la diferenciación de cada uno de los métodos dentro de modelización de tópicos, se va a analizar a fondo el Modelo de Localización Latente de Dirichlet.

Ilustración 2: Conceptualización del Modelo Localización Latente de Dirichlet



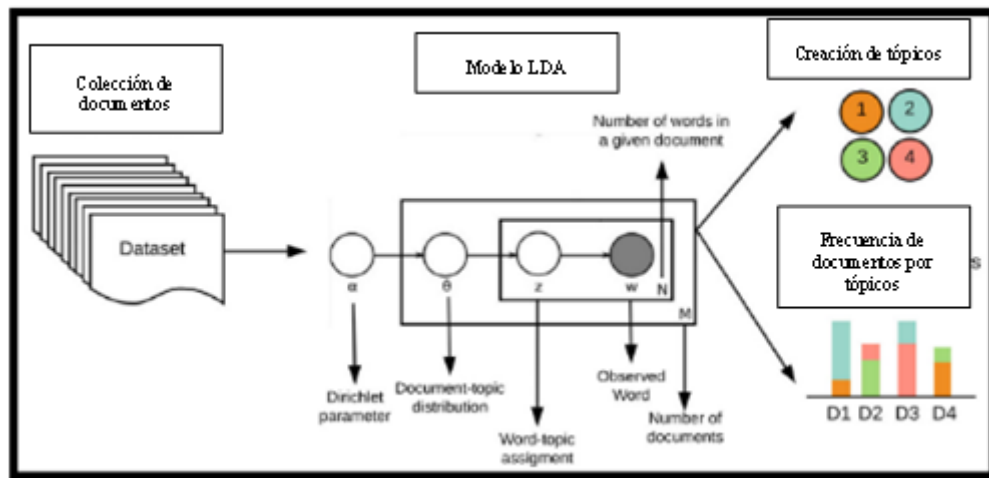
Fuente: (Bechmann y Bowker, 2019)

Nota: Proceso de la Modelización de tópicos por localización Latente de Dirichlet, extracción de palabras claves y creación de tópicos

2.2.4. Localización Latente de Dirichlet.

Este modelo, fue introducido por primera vez por David M. Blei y Andrew Y. Ng en 2003, quienes lo definieron como “Técnica de análisis computacional que se puede utilizar para investigar la estructura temática de una colección de datos escritos” (Buenano-Fernandez et al., 2020). Este algoritmo combina la técnica inductiva con herramientas estadísticas, lo que lo convierte en una técnica adecuada para la exploración y el análisis descriptivo (Pietsch & Lessmann, 2018).

Ilustración 3: Esquemización del Modelo Localización Latente de Dirichlet(LDA)



Fuente: (Buenano-Fernandez et al., 2020).

Esquemización del algoritmo de Dirichlet, desde la colección de los documentos de textos, aplicación de la distribución de Dirichlet, creación de tópicos y frecuencia de tópicos por documentos. (Chauhan & Shah, 2021).

(Echeverría-Villafuerte y Villarruel-Meythaler, 2022) proponen algunos pasos que debe seguir el modelo de Localización Latente de Dirichlet. Primero se seleccionan tópicos predefinidos, y las palabras más probables de aparecer en estos tópicos, segundo, escoger un conjunto de tópicos para cada palabra y una palabra para cada tópico, y tercero, se agrupan los tópicos relacionados a las palabras que aparecen con mayor frecuencia en el texto. Esto nos ayuda a identificar los tópicos que mejor explican los datos del corpus.

2.2.5. Aplicaciones del Modelo de Localización Latente de Dirichlet

La aplicación del Modelo de Localización Latente de Dirichlet es muy amplia, pues puede ir desde lo más sencillo como el análisis de preguntas abiertas de una encuesta hasta análisis más complicados como el análisis de documentos relacionados a las tecnologías, como también ser aplicado en diferentes contextos. Es así como, por ejemplo (Pietsch y Lessmann, 2018). Utiliza varios métodos dentro de la modelización de tópicos para

analizar una pregunta abierta de una encuesta realizada por una empresa en Berlín, donde se preguntaba a los desarrolladores por qué recomiendan desarrollar en una determinada plataforma, a estos datos se les realizó el procesamiento como traducción de palabras en inglés, lemanización, tokenización, conversión a minúsculas, eliminación de signos de puntuación, etc., con el fin de mejorar la calidad del conjunto de datos y modificarlos para que cumplan con los requisitos del modelamiento de tópicos (de texto corto), obteniendo como resultado que el la Localización Latente de Dirichlet es el modelo que más cubría con los objetivos del análisis, además señala que las preguntas abiertas son muy importantes para estudios sociales o de investigación de mercado, sin embargo, están asociadas a un análisis difícil por lo que es propenso a errores denominado codificación humana.

Uno de los contextos en los que se ha desarrollado este modelo ha sido en el análisis de redes sociales. (Bechmann y Bowker, 2019) enlistan varios casos de estudio donde la metodología empleada fue la modelización de tópicos, en uno de ellos se aplicó Localización Latente de Dirichlet para comprender el contenido a las que los usuarios están expuestos en el *feed* de noticias de Facebook, y entender los mensajes que hacen que una población específica se haga viral. Otro ejemplo aplicado al contexto de las redes sociales es el de (Diaz-Rubiano, 2022) quien realizó un estudio de caso aplicado al caso colombiano aplicando análisis de temas en la red social Twitter, el objetivo en este estudio fue analizar lo que los colombianos piensan y expresan en Twitter, llegando a la conclusión que los usuarios usan esta red social para quejarse sobre la coyuntura del país, por otro lado, se crearon temas como la afinación electoral, dividiéndose en dos, el grupo que prefería al neoliberalismo y otro grupo al socialismo, otra tema estuvo relacionado al

proceso de vacunación contra el COVID 19, el tercer tema tuvo relación con programas televisivos como Master Chef y los premios de MTV.

Otro contexto en el que se aplica Localización Latente de Dirichlet son las publicaciones de artículos en revistas o textos periodísticos disponibles en el internet. Por ejemplo, (Pietsch & Lessmann, 2018) y (Vernazza, 2020) analizan lo que se publica sobre el COVID 19, ambas autoras trabajan como base de datos a artículos científicos publicados. Por un lado (Pietsch y Lessmann, 2018) aplicaron este modelo a 4928 resúmenes científicos publicados en PubMed durante el primer semestre del 2020, obteniendo como resultados dos grandes tópicos, el primero corresponde a factores de riesgo, severidad y mortalidad por infección viral, el segundo tópico hace referencia al impacto de las infecciones respiratorias en la salud pública, la autora concluyó que entonces son dos líneas de investigación que toman los investigadores al hablar del COVID 19.

Por otro lado, este mismo autor utilizó 59887 artículos publicados entre 1951 y 2020 con temáticas asociadas a los diversos tipos de coronavirus, haciendo énfasis al COVID 19, las autoras realizaron un análisis descriptivo complementando su análisis con Asignación Latente de Dirichlet para la determinación de tópicos, sus resultados fueron que aquellos artículos publicados entre el 1 de enero y el 5 de mayo de 2020 pueden ser agrupados en dos grandes grupos o tópicos definidos por las palabras más utilizadas.

Otro ejemplo, dentro de este contexto es el estudio de Localización Latente de Dirichlet aplicada a textos periodísticos realizado por (AnalyStats, 2019) en este se realizó un análisis textual de los artículos de la sección de Finanzas de los diarios Valor Económico de Brasil y Portafolio de Colombia, después de la obtención de los datos, se realizó la depuración necesaria y se procedió a aplicar la función LDA del software R, una vez aplicada esta función, se obtuvieron 4 tópicos para cada diario periodístico. Para Valor

Económico se obtuvieron como tópicos a: Noticias sobre alguna caída influenciada por la Reserva Federal, el segundo tópico a noticias sobre algún evento sindical ocurrido en Argentina, el tercer tópico a noticias sobre el mercado financiero brasileño y el cuarto tópico a noticias sobre los resultados del Banco Santander. Para el diario Portafolio se obtuvieron los siguientes tópicos: el primero hace referencia a noticias sobre proyectos energéticos de Ecopetrol, el segundo a noticias sobre servicios o proyectos digitales o tecnológicos, el tercero, noticias sobre negocios de compañías en Bogotá y el cuarto, servicios de aerolíneas de Bogotá.

Este modelo también puede ser aplicado en el contexto empresarial, así lo ejemplifica el estudio de caso realizado por (Islam et al., 2022) quien realiza una modelización de los precios de Airbnb perteneciente al Condado de San José la cual ayuda a la toma de decisiones en cuanto a la aceptabilidad y rentabilidad del servicio. El autor indica que la descripción adecuada de los anuncios de Airbnb y su ubicación son determinantes para el precio. La modelización de tópicos dio como resultado dos tópicos, aquellos relacionados con los que determinan aquellos precios altos, los cuales intervienen: número de habitaciones, alojamiento, tipo de propiedad, número de reseñas. Mientras para los precios bajos están: Ausencia de insignia del super-anfitrión, política de cancelación. Estas conclusiones pueden ayudar a los propietarios a determinar precios de listado considerando la ubicación y los vecindarios circundantes.

Otro ejemplo similar es el de (Annisa & Surjandari, 2019) quienes aplicaron Localización Latente de Dirichlet para analizar la opción de los clientes aplicado al sector hotelero en Mandalika uno de los sitios turísticos de Indonesia. El objetivo de este estudio fue el de inferir los temas extraídos de las reseñas de hoteles de Mandalika. Para la obtención de los datos el autor aplicó *Web Scraping* a la página web de Traveloka obteniendo 1187

reseñas de los hoteles utilizando Python. Al aplicar Dirichlet se obtuvieron 8 tópicos: Satisfacción con la habitación, Instalaciones, desayuno, ubicación del hotel hacia la playa, experiencia turística alrededor del hotel, ambiente, hospitalidad y baño. El resultado de este estudio sirvió como retroalimentación a la gestión hotelera con el fin de aumentar el número de turistas que visitan Mandalika y la mejora de la satisfacción de ellos. Por su parte, (Slof et al., 2021) realizó un modelo de riesgos competitivos con Localización Latente de Dirichlet para predecir las razones de abandono de clientes de una empresa proveedora de servicios de telecomunicaciones, su objetivo radicaba en comprender el comportamiento de abandono de los clientes para reducir esta tasa, mediante este método se logró determinar 4 tópicos, cada uno relacionado a un tipo de riesgo, cada uno con palabras específicas: Riesgo de abandono, Riesgo controlable, Riesgo incontrolable y riesgo desconocido, ayudando al área de marketing a establecer estrategias para evitar que las tasas de abandono de los clientes sean altas o se incrementen, comprendiendo las necesidades del cliente a profundidad.

Por último, en el contexto de educación se han encontrado estudios de casos como el de (Echeverría-Villafuerte y Villarruel-Meythaler, 2022) quienes describieron el perfil de egreso de los estudiantes de la carrera de Estadística de la Universidad Central del Ecuador mediante la malla curricular de esta carrera, además con ayuda de la técnica *Web Scraping* y Localización Latente de Dirichlet describieron los perfiles profesionales que requieren las empresas en el portal de empleo "Opción empleo", obteniendo como resultados: Para el perfil de egreso de la carrera de Estadística se obtuvieron los siguientes tópicos: Aspectos pedagógicos, uso de software, conocimientos circunscritos y por último conocimientos específicos. En cambio, en el análisis del perfil profesional en Opción empleo se hallaron 25 tópicos, de los que se resaltan a competencias centradas en analítica

avanzada, ingeniería y arquitectura de datos, Inteligencia de negocios, formación académica, experiencia e idiomas, Machine Learning y Big Data, sectores como finanzas, seguros, salud y docencia. Otro caso aplicado para este contexto es el de (Buenano-Fernandez et al., 2020) quien aplicó Modelización de Tópicos por Localización Latente de Dirichlet para analizar las preguntas abiertas en la autoevaluación de profesores universitarios en una universidad ecuatoriana, el autor lo realizó en 4 pasos, el primero, construcción de la base de datos de texto, segundo, aplicación de minería de texto y modelado de temas, el tercero, modelado de redes de temas y cuarto la relevancia de los temas identificados. Con ayuda de esta metodología se hallaron 12 tópicos, resumidos en la siguiente ilustración:

Ilustración 4: Tópicos creados por LDA para el estudio de caso en el contexto educativo

Topic	Description
Topic 1	Research, analysis and reading
Topic 2	No definition
Topic 3	Tutorship
Topic 4	Use of technology
Topic 5	No definition
Topic 6	Practical learning
Topic 7	Practical learning
Topic 8	Retention strategies
Topic 9	Teaching-learning environment
Topic 10	Experiential learning strategy
Topic 11	Evaluation mechanisms
Topic 12	Team work

Fuente: (Buenano-Fernandez et al., 2020)

Mediante la agrupación de la red de tópicos se seleccionaron 4 grupos o clústeres.

- Clúster 1 (amarillo) Aprendizaje practico: Incluye los temas 5, 6 y 7
- Clúster 2 (Azul) Iniciativa de enseñanza: Incluyen temas 8, 9 y 10
- Clúster 3 (Rojo) Uso de herramientas tecnológicas y estratégicas de enseñanzas tradicionales: Tema 1, 4, 11

- Clúster 4 (Verde): Estrategias de trabajo en equipo: tema 12

El autor llegó a la conclusión que es importante dar un análisis profundo a las preguntas abiertas de las encuestas de calificación a docentes universitarios, pues estas indican el grado de conformidad de los estudiantes en cuanto a la educación y con ella la calidad de esta. Además, resalta la importancia de aplicar técnicas profundas no clásicas para el análisis de texto.

2.3. FUNDAMENTACIÓN LEGAL

2.3.1. Aspectos Legales del *Web Scraping* en el Ecuador.

En la actualidad no existe leyes o normas que regulen la *Web Scraping* como tal, esta técnica está incluida a otros aspectos legales como: los derechos de propiedad intelectual, es decir no se puede tomar otra información ajena como si fuera propia, se deben respetar los términos de servicio de los sitios web, y los datos no deben ser objeto de una competencia desleal (Ramírez, 2022).

2.3.2. Ley de Comercio Electrónico, Firmas y Mensajes de Datos – 2002:

Artículo 9 Sección I: La recopilación y el uso de los datos personales será realizado respetando el derecho a la privacidad, intimidad y confidencialidad (Ley de comercio electrónico, firmas y mensajes de datos, 2002).

2.3.3. Ley Orgánica de Telecomunicaciones - 2015

Artículo 22 Capítulo I: Exige a los prestadores de servicios proteger los datos de sus usuarios, sin embargo, no se da una instrucción clara de cómo hacerlo. Se hace énfasis en que los usuarios tienen el derecho de que sus derechos sean protegidos y que las empresas tienen la obligación de hacer lo necesario para que los datos de sus usuarios no sean violentados (Ley Orgánica de Telecomunicaciones, 2015).

2.3.4. Ley Orgánica de Protección de Datos Personales (LOPDP) – 2021

En esta, se establece un marco normativo completo sobre los derechos que tienen los individuos en la protección de sus datos.

Artículo 12: Se establece el derecho a la información, como un principio de transparencia, en donde el usuario tiene derecho a saber el fin de sus datos, para qué son utilizados y en qué escenarios, sobre todo, con estar de acuerdo con que sus datos sean utilizados, y no debe ser de desconocimiento lo que se hace con ellos (Ley Orgánica de Protección de Datos Personales, 2021).

Es importante señalar que para la aplicación de la técnica de *Web Scraping* dentro del portal de empleo Hiring Room, la obtención de la información valiosa se encuentra totalmente anonimizada, en otras palabras, no se toman los datos de personas específicas, violentando su derecho a la privación de sus datos. De esta forma no se viola la LOPDP, respetando los datos personales y la libre circulación de estos.

CAPÍTULO III

METODOLOGÍA DE LA INVESTIGACIÓN

3.1. TIPO DE INVESTIGACIÓN

Principalmente, el tipo de investigación que se abordó en este trabajo de investigación es de tipo descriptivo. Pues se describió características o fenómenos particulares en detalle. Para este análisis se emplearon técnicas que describieron de forma cualitativa y cuantitativa los perfiles de egreso de los profesionales de Telemática y Software de la Universidad Técnica Estatal de Quevedo, como también se describieron los perfiles que demandan las empresas en los portales de empleo.

3.2. MÉTODOS UTILIZADOS EN LA INVESTIGACIÓN

3.2.1. Método Analítico:

Se realizó un análisis lógico y conceptual del tema, para comprender las relaciones lógicas y teóricas aplicadas a la problemática, con el fin de entender qué relaciones existen entre ellas explicándolas de forma efectiva, mediante una revisión bibliográfica profunda.

3.2.2. Método deductivo:

Mediante las aplicaciones de teorías y teoremas como el Teorema de Bayes, Modelización de tópicos, entre otros, se analizó un fenómeno específico como lo es la aplicación de Localización Latente de Dirichlet en la caracterización del perfil de egreso de estudiantes y el perfil de profesionales demandados.

3.2.3. Método Estadístico:

Se ha aplicado este método desde la obtención de los datos hasta la representación de resultados, con el fin de obtener conclusiones validas, explicados a continuación.

1. Se recopiló la información por medio de la técnica *Web Scraping* en los portales de empleo.

2. Se analizó la información mediante la aplicación del método estadístico de Localización Latente de Dirichlet, Minería de Texto y creación de tópicos en la data obtenida por los portales de empleo como en las mallas curriculares de las carreras de la facultad de Ciencia de la Ingeniería.

3. Representación de los resultados.

4. Conclusiones, al explicar y analizar la brecha encontrada.

3.3. CONSTRUCCIÓN METODOLÓGICA DEL OBJETO DE INVESTIGACIÓN.

3.3.1. Población y muestra

3.3.1.1. Población

El análisis de nuestra población se dividió en dos partes tanto para los anuncios de empleo como las capacidades específicas de los programas analíticos de las carreras.

Anuncios publicados por las empresas en los portales de empleo:

La población por esta parte fueron los anuncios publicados por empresas en Hiring Room en el último año a nivel nacional.

- Unidad de análisis: anuncios de empleo publicados.

Capacidades específicas y perfil de egreso de los profesionales graduados en el Universidad Técnica Estatal de Quevedo

La población en este caso, fueron las mallas curriculares y planes analíticos del segundo periodo 2023 –2024, se seleccionaron estas fuentes de información puesto que, estos documentos detallan de forma clara y general los temas, subtemas, objetivos y resultados de cada asignatura de las carreras de Software y Telemática durante todo el semestre.

- Unidad de análisis: Malla curricular y planes analíticos, en el Anexo 15 se puede observar un Plan Analítico de la carrera de Software, su estructura y los campos que fueron extraídos.

3.3.1.2. Muestra

Anuncios publicados por las empresas en los portales de empleo:

Para el cálculo de la muestra se utilizó la fórmula del tamaño de la muestra para una población infinita (o población desconocida). En el Anexo 14 se puede observar el ejemplo de un anuncio de empleo, su estructura y los datos que fueron extraídos.

Ecuación 2: Calculo del tamaño de la muestra de anuncios publicados en Hiring Room

$$n = \frac{Z_{\alpha}^2 \times p \times q}{e^2}$$

Donde:

n = es el tamaño de muestra buscado

Z = es el parámetro estadístico, el cual depende del Nivel de Confianza

NC = nivel de confianza al 95%

e = error de estimación máximo aceptado 5%

p = probabilidad de éxito

q = probabilidad de no éxito

$$n = \frac{(1.96)^2 \times 0.5 \times 0.5}{0.05^2}$$

$$n = 356$$

Por lo tanto, se trabajó con 356 publicaciones de empleo en el periodo 2022-2023, escogidas al azar, siempre y cuando tengan relación con la rama de profesionales escogidos.

Capacidades específicas y perfil de egreso de los profesionales graduados en el Universidad Técnica Estatal de Quevedo

Para el análisis de las capacidades específicas y generales de los estudiantes de las carreras de Software y Telemática se escogieron aquellas asignaturas de formación profesional por conveniencia del investigador

3.3.2 Técnicas de investigación

Entre las técnicas de investigación utilizadas en el presente proyecto de investigación están:

Técnica de análisis documental y revisión bibliográfica:

Con esta técnica se escogieron y evaluaron documentos, textos y materiales relevantes que explicaron y sustentaron teóricamente el tema de investigación, como también a la metodología que se empleó. Esta técnica permitió además dar un marco teórico y un estado del arte que defina hasta qué punto se debe analizar la información.

Técnica de investigación en línea y Big data:

La producción masiva de los datos y el auge de las técnicas de procesamiento de datos, han desarrollado técnicas para analizar datos en línea, como el análisis de redes sociales, minería de datos y texto, inteligencia artificial, etc. En este proyecto de investigación se empleó profundamente esta técnica de investigación para obtener datos específicos e implementar modelos para el análisis de lenguaje natural.

Técnicas de minería de texto:

Se aplicó modelos de Modelización de Tópicos como también funciones de visualización de texto y minería de texto como *Wordcloud*.

3.3.3 Instrumentos de Investigación

Banco de datos – Anuncios de empleo:

Se obtiene una base de datos para recolectar datos secundarios previamente generados por un organismo externo como entidades gubernamentales o portales web. En nuestro caso, usamos este instrumento para la obtención de una data específica partiendo de datos estructurados anteriormente en un portal de empleo.

Paquetes y bibliotecas de programación de R

El lenguaje de programación R fue un instrumento clave para llevar a cabo todos los objetivos del trabajo de investigación mediante este se realizó *Web Scraping*, aplicación del modelo de Modalización de Tópicos y el análisis de la brecha con minería de texto.

3.4. ELABORACIÓN DEL MARCO TEÓRICO

Para la elaboración del marco teórico se realizó de forma sistemática con los siguientes pasos: Primero, se definió el tema y lo objetivos que se quieren lograr para responder a una problemática específica. Segundo, se desarrolló una revisión sobre lo métodos que se aplicaron. Métodos de extracción de datos, tipos de metodologías que se usaron para alcanzar los objetivos. Tercero, se realizó una búsqueda exhaustiva de literatura que den soporte teórico a la problemática. Dentro de este paso, se tomó en cuenta aquella literatura publicada en repositorios de universidades (Tesis), revistas científicas, portales web altamente calificados, videos y conferencias, todas las fuentes con una antigüedad

máxima de 5 años cuyo idioma pudo ser español e inglés, y sin importar la región en la que fue realizada.

Otro criterio para seleccionar las fuentes bibliográficas fue que estas publicaciones deben contar con palabras claves como: ciencia de datos, minería de texto, modelización de texto, Localización Latente de Dirichlet, mallas curriculares, educación universitaria, brecha laboral, mercado laboral. Cuarto, se seleccionaron los estudios más relacionados a la temática o problemática, en este paso se seleccionaron los estudios relacionados con técnicas y metodologías semejantes a las de nuestros objetivos. Quinto y, por último, se creó el marco teórico que sintetizó el criterio de otros escritores en conjunto con el criterio propio del autor.

Para la fundamentación conceptual se definieron aquellos términos que son necesarios comprender para que la comprensión de la temática y contextualización de la problemática. Para la fundamentación teórica, se citaron de forma lógica todas aquellas investigaciones relacionadas al tema en cuestión, en especial la modelización de tópicos con Localización Latente de Dirichlet. Por último, para la fundamentación legal, se citaron aquellas normas, y leyes relacionadas al manejo de información personal, que se rigen actualmente en el Ecuador, vea anexo 2,

3.5 RECOLECCIÓN DE LA INFORMACIÓN

La obtención de los datos se realizó en dos secciones.

1. Portal de empleo Hiring Room

La recolección de los datos en esta sección se la realizó por medio de la técnica *Web Scraping*, es decir, los datos que se encuentran publicados en la página web Hiring Room en forma de anuncio de empleo fueron descargados y almacenados en una base de datos no estructurada. Se extrajeron variables tales como:

- Nombre de la empresa
- Nombre de la vacante
- Descripción del puesto
- Ubicación
- Áreas
- Requisitos educación
- Requisitos experiencia
- Requisitos conocimientos
- Beneficios

Ver anexo 14, para observar la estructura y los datos que fueron extraídos.

2. Mallas curriculares y Planes Analíticos

En esta sección se recolectaron las mallas curriculares y planes analíticos de las carreras de Software y Telemática de los semestres superiores a sexto a décimo (de formación profesional. Utilizando la técnica *Web Scraping* se extrajeron datos como:

- Número de créditos, horas, semestres
- Materias teóricas
- Materias profesionales
- Nombre de la asignatura
- Objetivos y resultados
- Temas y subtemas

- Metodología

Ver Anexo 15 para observar la estructura de un plan analítico y los datos a extraer.

3.6 PROCESAMIENTO Y ANÁLISIS

3.6.1. Perfil demandando por las empresas en el portal de empleo Hiring Room

Para la obtención de los datos de los anuncios de empleo relacionados al área de las Tecnología de la Información y Comunicación, se aplicó la técnica de *Web Scraping* a los anuncios de empleo mediante la librería Rvest del lenguaje de programación R-studio con todos los enlaces seleccionados del portal de empleo, se identificaron los selectores de los elementos a extraer, en este caso el nombre de la vacante, ubicación, empresa, requisitos. Se extrajeron los datos seleccionados por medio de la función `html_text` o `html_nodes`. Los datos fueron guardados en una base de datos de Excel. Una vez obtenida la base de datos, se analizaron los datos mediante herramientas de estadísticas descriptivas como gráficos y tablas de frecuencias.

Para la aplicación de la Modelización de Tópicos por Localización Latente de Dirichlet, se aplicaron técnicas de preprocesamiento a la descripción, requisitos de educación, experiencia y requisitos de conocimientos de las vacantes, como lemanización, toquenización, transformación de mayúsculas a minúsculas, eliminación de signos de puntuación y otros. Se transformaron las descripciones en un objeto de tipo Corpus, usando la función `corpus` de la librería `tm`, Luego, se creó una Matriz de Términos de Documento usando la función `DocumentTermMatrix()` la cual almacenó la frecuencia de términos en cada documento. Para finalmente definir la cantidad de tópicos y entrenar el Modelo de Localización Latente de Dirichlet. Para su análisis fue necesario ver los

términos más representativos por cada tópico, y de igual forma para cada documento y dar un nombre a cada tópico encontrado, y así encontrar el perfil del candidato ideal.

3.6.2. Descripción del perfil de egreso de los estudiantes de las carreras de Software y Telemática

Para la creación de la base de datos del perfil de egreso de los estudiantes se utilizó la técnica de *web scraping* a los planes analíticos vigentes de las carreras de Software y Telemática utilizando la librería *pdftools* del lenguaje de programación R-studio, extrayendo información como nombre de la asignatura, cantidad de créditos, cantidad de horas, prerrequisitos, descripción, objetivos y resultados de las asignaturas, los cuales se almacenaron en una base de datos en Excel. Una vez obtenida los datos, se procedió a realizar técnicas de preprocesamiento de texto y aplicación de la Localización Latente de Dirichlet, para hallar tópicos conformada por términos y documentos. Una vez hallados los tópicos se crearon los perfiles de los egresados. Además, se realizó un análisis descriptivo de las mallas curriculares de ambas carreras para analizar datos como cantidad de horas, créditos, semestres que tiene cada una de ellas.

3.6.3. Analizar la brecha existente entre los tópicos identificados con el modelo de Localización Latente de DirichLet en las competencias específicas y los anuncios de empleo, mediante técnicas de minería de texto

Una vez descritos y hallados ambos perfiles se procedió a analizar la brecha existente entre estos, mediante herramientas de minería de texto como lo son las funciones *commonality* y *comparison*, de las librerías *tm* y *wordcloud* del lenguaje de programación R-studio las cuales permitieron encontrar las palabras comunes entre el perfil de egreso y el perfil demandado por las empresas, y comparar las palabras para cada uno de los perfiles demostrando la brecha entre ellos.

CAPÍTULO IV

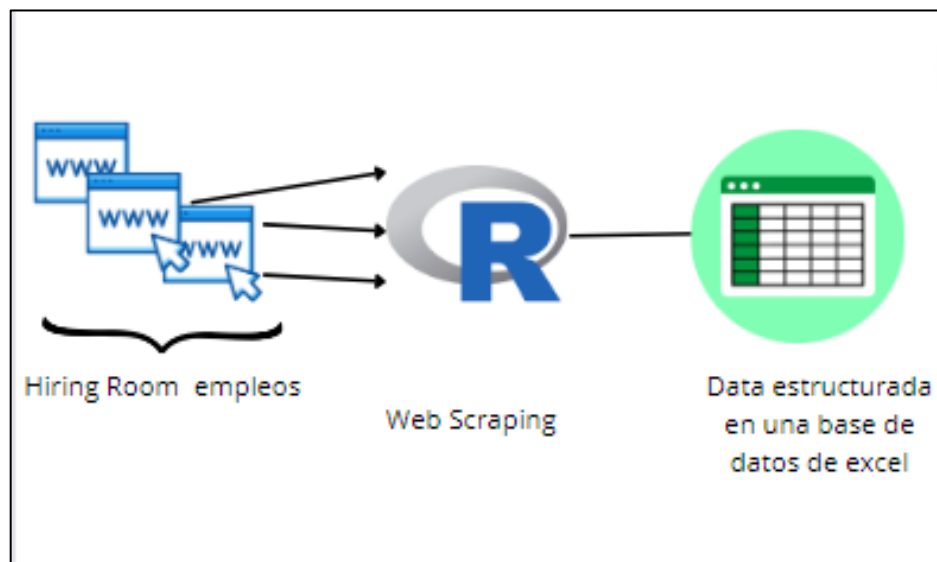
RESULTADOS Y DISCUSIÓN

4.1.APLICACIÓN DE LAS TÉCNICAS *WEB SCRAPING* Y MODELIZACIÓN DE TÓPICOS A LOS ANUNCIOS DE EMPLEO DEL PORTAL HIRING ROOM, PARA LA DESCRIPCIÓN DE LOS PERFILES DE VACANTES DEMANDADAS POR LAS EMPRESAS PÚBLICAS Y PRIVADAS

4.1.1. Aplicación de *Web scraping* a los anuncios de empleo del portal Hiring Room

Para la aplicación del *web scraping*, se necesitó contar con los enlaces de cada anuncio de empleo, para la posterior selección de los datos a extraer y finalmente el almacenamiento de estos datos. A continuación, en la ilustración 5 se visualiza el procedimiento de esta técnica hasta la creación de la base de datos

Ilustración 5: *Proceso de la extracción de los anuncios de empleo mediante Web Scraping*

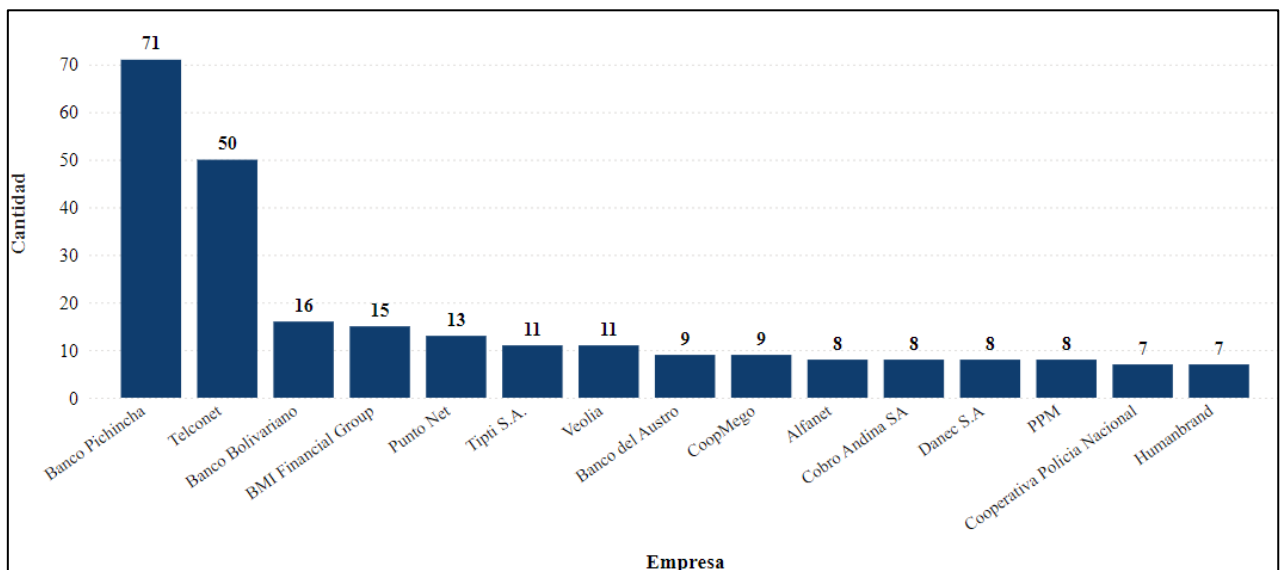


Fuente: Autor

Una vez creada la base de datos de los anuncios de empleo, se ejecutaron técnicas de preprocesamiento de texto como: Eliminación de signos de puntuación, artículos, preposiciones, cambio de palabras mayúsculas en minúsculas, lemanización, tokenización, entre otras.

Después, se analizó la información que dan una idea sobre las características de los datos. A continuación, se detallan gráficos y tablas que se obtuvieron aplicando estadísticas descriptivas.

Ilustración 6: *Top 15 de las empresas que más anuncios de empleo publican en Hiring Room en la búsqueda de profesionales del área de las TIC's*



Se encontraron 55 empresas a nivel nacional que publicaron anuncios de empleo en el portal Hiring Room en búsqueda de profesionales dentro del área de las TIC's, sin embargo, entre las primeras 15 se encuentran las siguientes ordenadas de forma descendente: Banco Pichincha la cual ha realizado 71 anuncios de empleo en esta área en los últimos 5 meses. Telconet con 50 anuncios de empleo, Banco Bolivariano con 16, BMI Financial Group con 15, Punto Net con 13, Tipti S.A. con 11 al igual que Veolia, Banco del Austro con 9, Alfanet con, Cobro Andina S.A, CoopMego, PPM con 8, Cooperativa Policía Nacional y Humanbrand con 7, los anuncios de estas 15 empresas representan el 70% de todos los anuncios.

En cuanto a la ubicación geográfica de estas empresas se tiene que:

Tabla 1: Ciudades en las que se ubican las empresas que publican anuncios de empleo

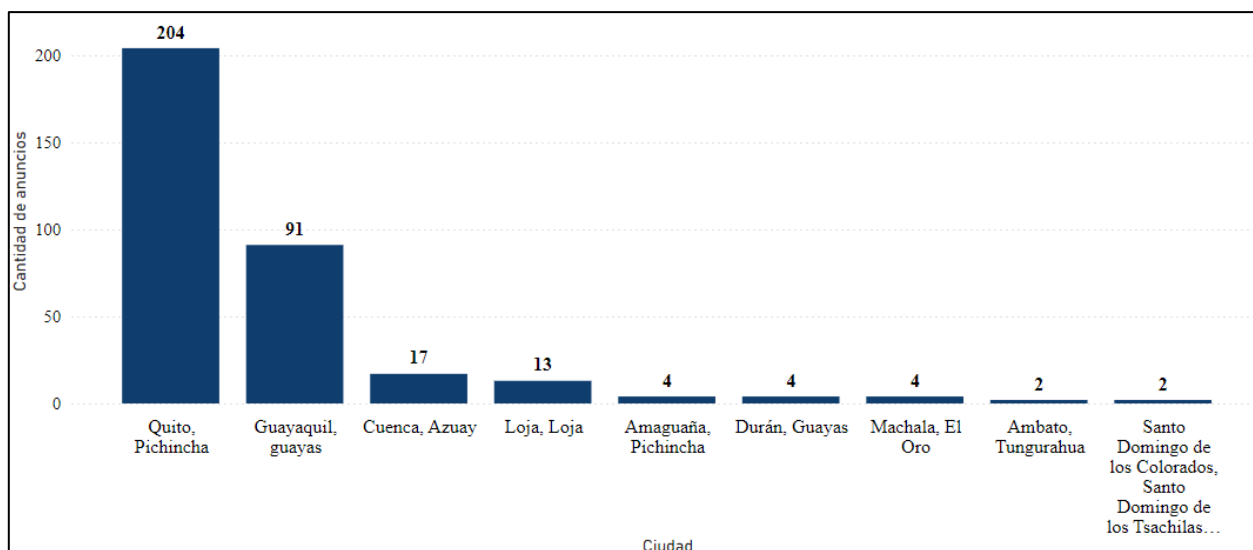
	Cantidad	Porcentaje	Porcentaje acumulado
Quito, Pichincha	204	57,1	57,1
Guayaquil, Guayas	91	25,5	82,6
Cuenca, Azuay	17	4,8	87,4
Loja, Loja	13	3,6	91,0
Amaguaña, Pichincha	4	1,1	92,2
Durán, Guayas	4	1,1	93,3
Machala, El Oro	4	1,1	94,4
Ambato, Tungurahua	2	0,6	95,0
S.D de los Colorados, S.D de los Tsáchilas,	2	0,6	95,5
Azogues, Cañar	1	0,3	96,1
Daule, Guayas	1	0,3	96,4
Gualaquiza, Morona Santiago	1	0,3	96,6
Guayllabamba, Pichincha	1	0,3	96,9
Pangui, Zamora Chinchipe	1	0,3	97,2
Pedernales, Manabí	1	0,3	97,5
Puembo, Pichincha	1	0,3	97,8
Quinindé, Esmeraldas	1	0,3	98,0
Riobamba, Chimborazo	1	0,3	98,3
Salinas, Santa Elena	1	0,3	98,6
Santa Isabel, Azuay	1	0,3	98,9
Sucúa, Morona Santiago	1	0,3	99,2
Tena, Napo	1	0,3	99,4
Zamora, Zamora Chinchipe	1	0,3	99,7

Zumba, Chinchi	Zamora	1	0,3	100,0
<hr/>				
Total		357	100,0	

Se encontraron 26 ciudades a nivel nacional, donde requieren personal en el área de las TIC's ya sea para trabajar en la Sede principal de la empresa o en alguna de sus sucursales, de la cual tenemos que la ciudad que atraen más profesionales del área de las TIC's es Quito con el 57,1% de todos los anuncios, Guayaquil con el 25,5%, Cuenca con el 4.8%, Loja con el 3,6%, Amaguaña, Duran y Machala con el 1,1%, Ambato y Santo Domingo con el 0.6%, y las otras 15 ciudades las cuales representan el 4,5% del total de empleo.

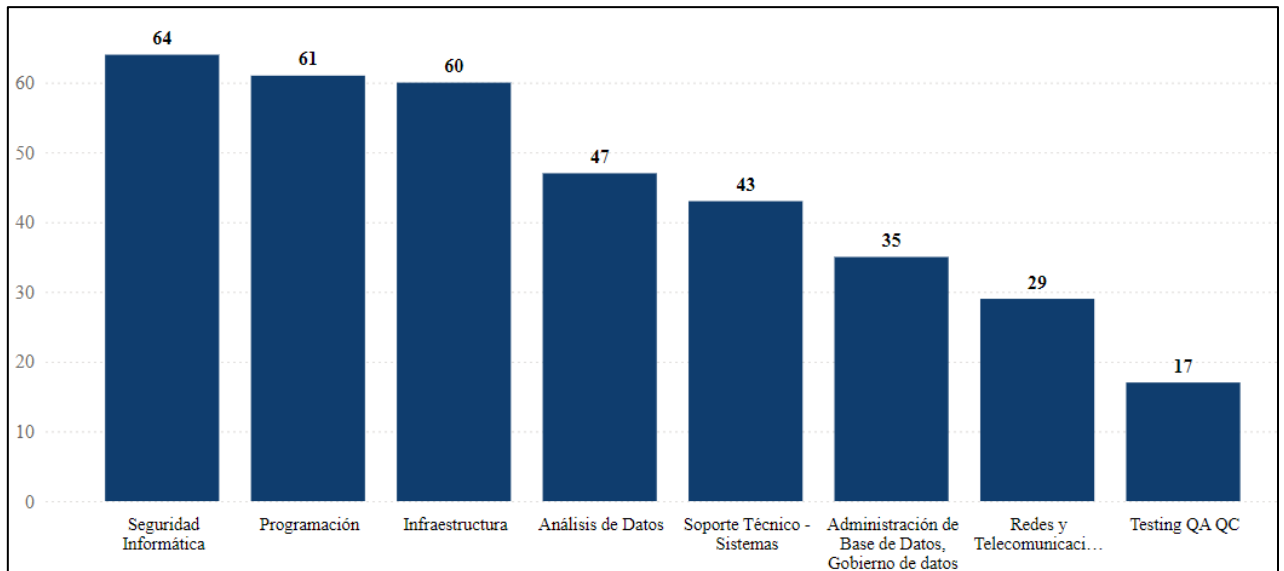
Solo entre Quito, Guayaquil, Cuenca, Loja, Amaguaña, Duran, Machala, Ambato y Santo Domingo se encuentra el 95.5% de los anuncios de empleo. A continuación, se puede observar un gráfico de barras que muestra esta información.

Ilustración 7: Ciudades a realizar las actividades de acuerdo con los anuncios de empleo



En cuanto a las áreas y subáreas de los anuncios de empleos, estos se distribuyen de la siguiente manera.

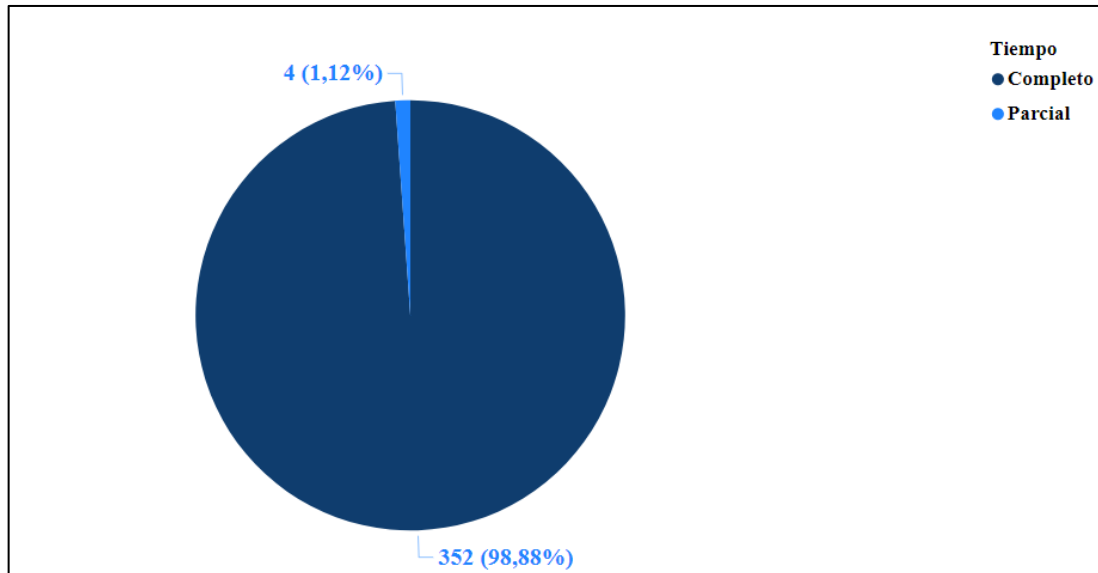
Ilustración 8: Anuncios de empleo de acuerdo con las subáreas de Tecnología, Sistemas y Telecomunicaciones



Se han detectado 8 sub-áreas de Tecnología, Sistemas y Telecomunicaciones, de las cuáles los perfiles con formación en Seguridad Informática es la más demandada con el 17,9% del total de anuncios de empleo, seguido de perfiles con alto componente en Desarrollo y Programación con el 17,1%, como tercer lugar está el perfil con formación en Infraestructura con el 16,8%, siguiendo Redes y Telecomunicaciones con el 8,1% y por ultimo aquellos perfiles con un alto conocimiento en Testing QA QC con el 5,1% de todos los anuncios de empleo.

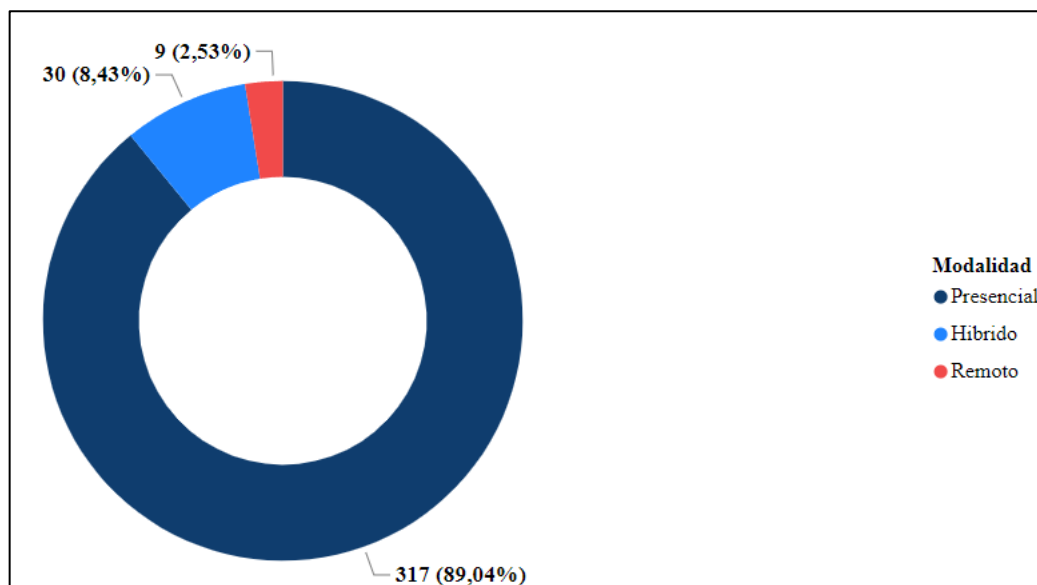
Todas estas actividades se realizan en dos jornadas o tiempo de trabajo, tiempo parcial y tiempo completo

Ilustración 9: *Tiempo de la Jornada Laboral*



El 98.88% de los anuncios de empleo indican que la jornada laboral es de 8 horas al día, mientras que el tiempo Parcial, medio tiempo indica que la jornada puede ser menor o igual a 4 horas

Ilustración 10: *Modalidad de trabajo*



El 89% de los anuncios de empleo de las empresas mencionadas son en la modalidad presencial, resulta interesante resaltar que existe el 8.4% de anuncios con modalidad híbrida, modalidad la cual se ha adoptado a partir de la pandemia por COVID 19. Por último, la modalidad con menos frecuencia, con el 2.5% en modalidad remota.

Tabla 2: *Modalidad de trabajo de acuerdo con la subárea de trabajo*

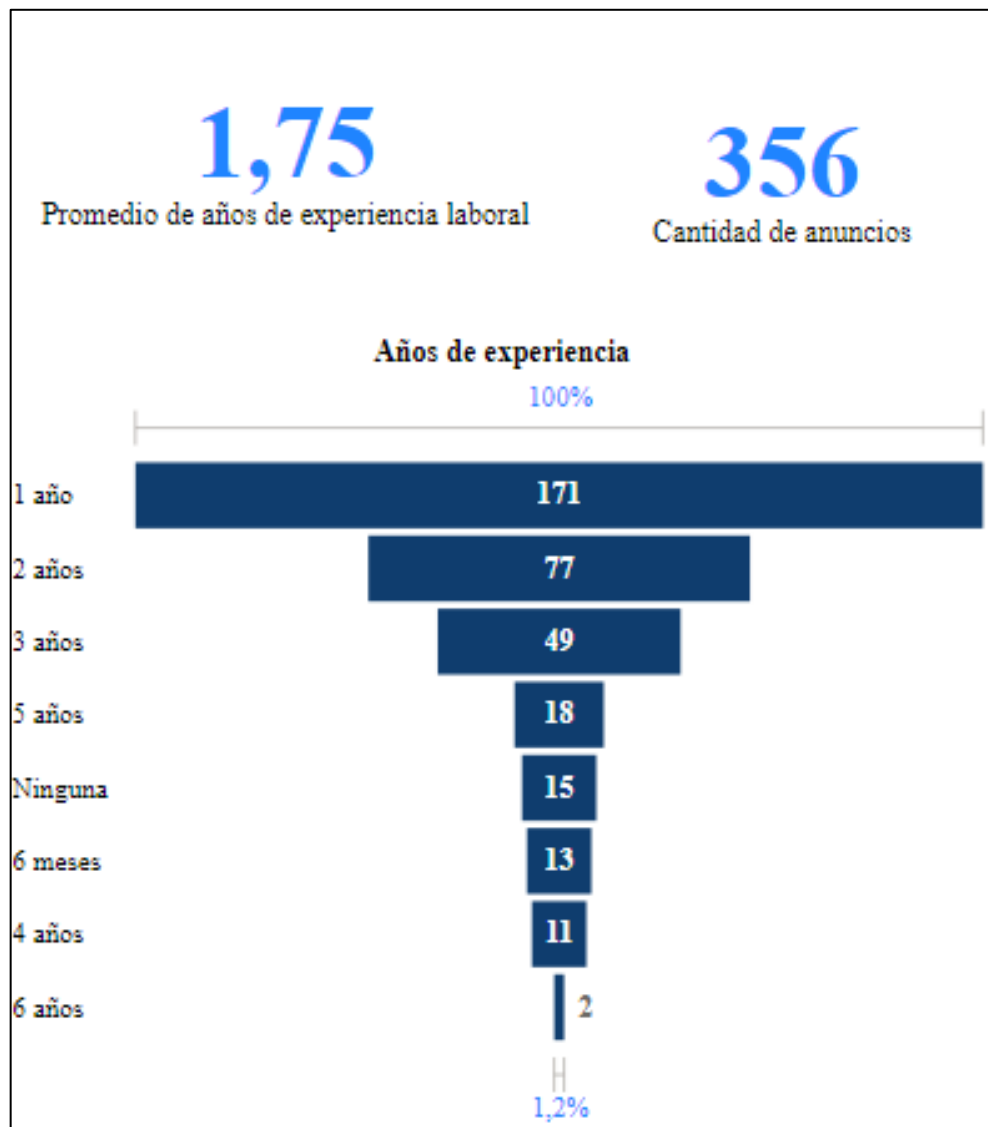
Sub-Área	Modalidad			Total
	Híbrido	Presencial	Remoto	
Administración de Base de Datos, Gobierno de datos	7	27	1	35
Análisis de Datos	6	41	0	47
Infraestructura	1	58	1	60
Programación	2	57	2	61
Redes y Telecomunicaciones	2	27	0	29
Seguridad Informática	4	57	3	64
Soporte Técnico – Sistemas	3	39	1	43
Testing QA QC	5	11	1	17
	30	317	9	356

En lo que respecta a la modalidad presencial, el 60% destacan en áreas como Infraestructura, Programación, Seguridad informática y análisis de datos. En el caso de la modalidad híbrida, en cual se desarrollan las actividades de forma presencial como de forma remota el 73.33% corresponden a áreas de Administración de base de datos,

Análisis de datos, Seguridad informática y Testing QA QC. Por último, en la modalidad remota, las áreas que destacan son Seguridad Informática y Programación.

Algo que es relevante al hablar del mercado laboral, es la cantidad de años de experiencia laboral, pues este es un factor que influye en la probabilidad de ser o no contratado.

Ilustración 11: *Años de experiencia laboral requeridos*



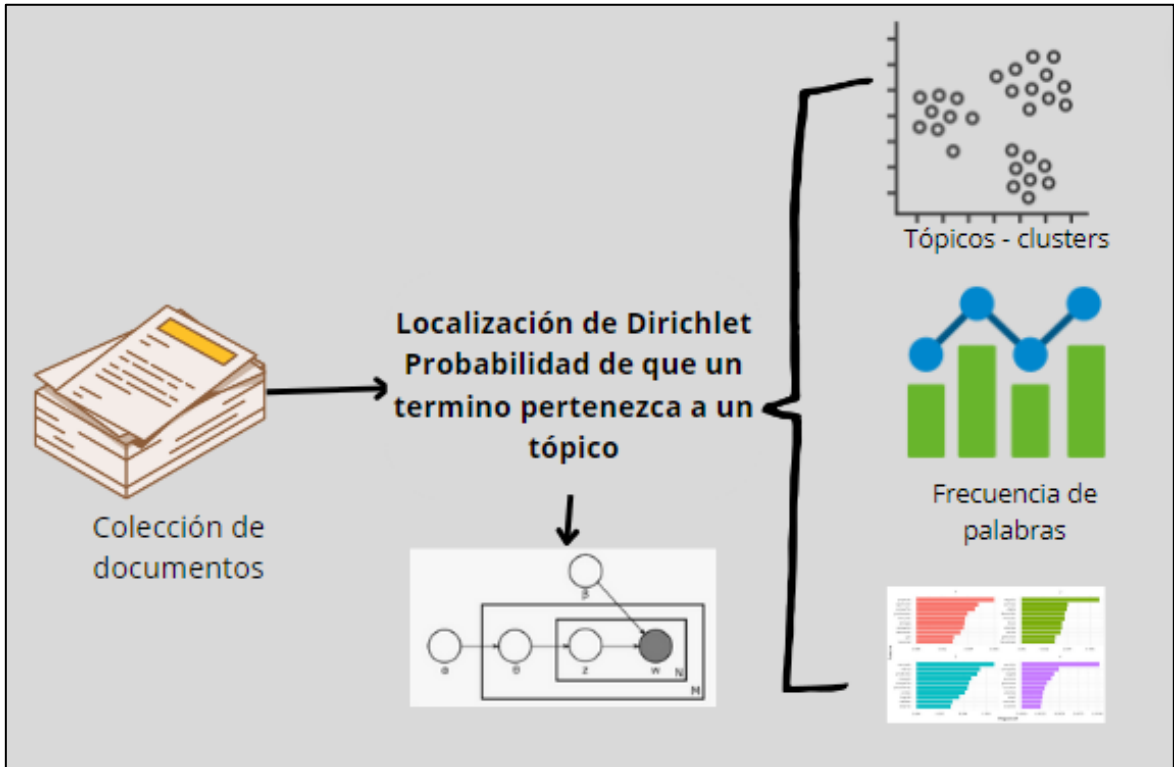
En cuanto a la experiencia laboral requerida, el 4.2% de las empresas no exigen años de experiencia laboral, más aún cuando se trata de prácticas profesionales, el 3.7% de las empresas requieren una experiencia mínima de 6 meses, en especial para aquellos puestos

Junior o asistentes, el 48% de las empresas exigen contar con 1 año de experiencia laboral, el 21.6% 2 años, el 13.8% 3 años, 3.1% 4 años, 5.1% 5 años y tan solo el 0.6% de las empresas 6 años de experiencia laboral. Por otro lado, el promedio de años de experiencia laboral a nivel nacional es de 1.75 años.

4.1.2. Aplicación de la técnica Localización Latente de Dirichlet (LDA) para descripción del puesto, requisitos de educación, conocimiento, experiencia y competencias generales.

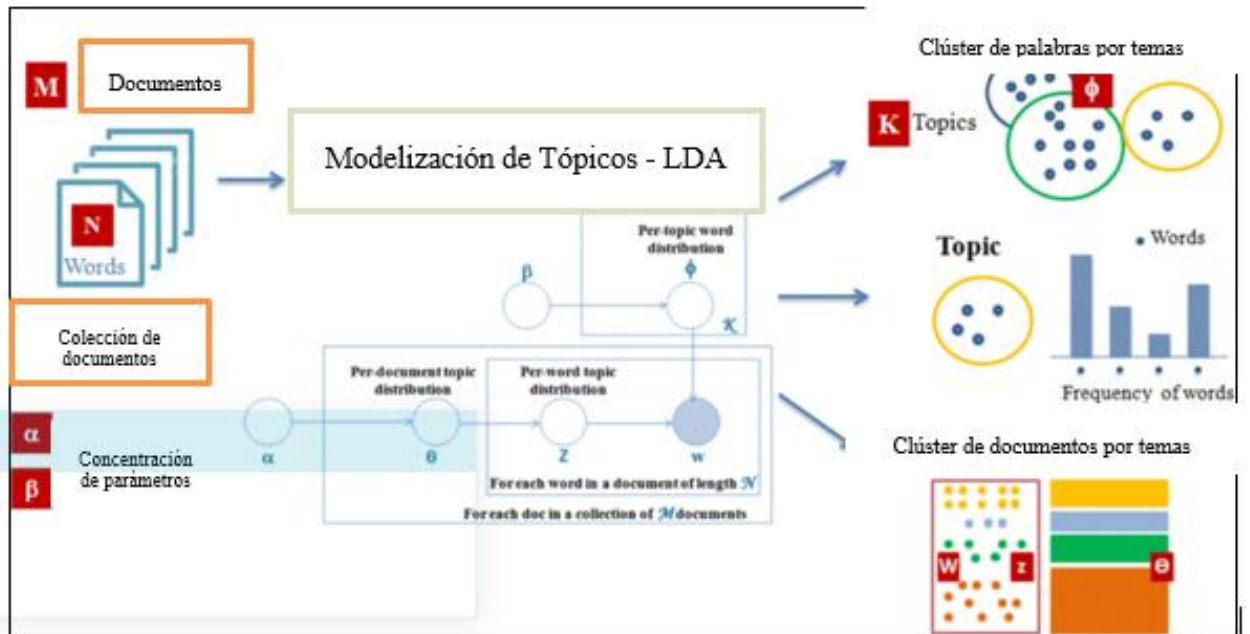
Para la obtención de los resultados obtenidos por el modelo de localización latente de Dirichlet, es necesario comprender paso a paso como se llegaron a estos. En las ilustraciones 12 y 13, se puede observar su procedimiento, en este caso un documento se encuentra representado por una vacante demandada en el portal Hiring Room, estos documentos, deben ser limpiados y revisados, es decir aplicar técnicas de preprocesamiento de texto, para luego ser transformados a un objeto de tipo corpus y a una matriz de términos, misma que es usada para hallar las probabilidades de que una palabra clave pertenezca a un tópico o tema específico o probabilidad Beta, de igual forma la probabilidad de que cada documento corresponda a un tópico o también conocido como probabilidad Gamma, tal como se puede observar en los anexos, donde se observa de forma gráfica las probabilidades de cada palabra clave pertenecer a un tópico.

Ilustración 12: Procedimiento lógico de la modelización de tópicos por Localización Latente de Dirichlet



Fuente: Autora

Ilustración 13: Procedimiento del algoritmo de Localización Latente de Dirichlet



Fuente: (Figure from *Latent Dirichlet Allocation (LDA) for Topic Modeling of the CFPB Consumer Complaints Semantic Scholar*, 2023)

En los anexos desde 5 al 8 se crearon múltiples tópicos con el fin de afinar el perfil del candidato ideal para las empresas resumiéndolos en la siguiente tabla.

Tabla 3: Perfiles hallados por Localización Latente de Dirichlet para las vacantes demandadas en el portal de empleo Hiring Room

Perfiles	Desarrollo de Software/ Aplicaciones	Ciberseguridad	Soporte técnico e infraestructura	Gestión y administración de bases de datos	Analítica de datos, modelos de Machine Learning	Redes y telecomunicaciones
Educación	Software Informáticos Ciencias computacionales Pasantas	Certificaciones ISO, CISSP, CISM, CEH, GCT Ingles Especializaciones en fraude o incidentes, cuarto nivel	Telecomunicaciones Telemática	Informáticos Software Cuarto nivel arquitectura de base de datos	Data Science, estadística, informáticos	Certificaciones en telecomunicaciones Telecomunicaciones Telemática Software Informáticos
Conocimientos	Desarrollo de software, Análisis de	Ciberseguridad, Auditoria y	soporte técnico programación y mesa y soporte	Arquitectura de datos, y redes, conocimiento en	Business intelligence, monitoreo y	Instalaciones de fibra óptica, Programación,

	calidad de software, aplicaciones web y móviles	fiscalización informática	técnicos, infraestructura	herramientas de la nube Gobierno de datos,	análisis de datos	electricidad, conductor
Habilidades blandas	Autodesarrollo, adaptación rápida, comunicación, funcional, resultados, atención al cliente, desarrollo en entornos digitales, resolución de problemas, asertivo, proactividad, gestión de las emociones, trabajo bajo presión, autodesarrollo.					

Discusión:

Los resultados obtenidos en la descripción de los anuncios de empleo como la elaboración de los perfiles de los profesionales indican la necesidad de la formación de profesionales interdisciplinarios, en áreas administrativas y financieras, de acuerdo con (Echeverría-Villafuerte & Villarruel-Meythaler, 2022), quienes de igual forma resaltaron la importancia de la formación de profesionales con conocimientos básicos a medios en otras disciplinas, adquiriendo de esta forma habilidades para poder competir en el mercado laboral. Se pudo observar además que los campos del conocimiento que sobresalen son: informática, matemática y administración, con conocimientos de ETL, Machine Learning, programación en lenguajes como Python, R, SPSS, SQL, Ciberseguridad, y la necesidad de contar con varias certificaciones o formación de cuarto nivel, lo que podría ser considerado como un factor importante en cuanto la brecha. Además, se encontró la demanda de ciertas habilidades blandas como responsabilidad, comunicación, trabajo bajo presión, atención al cliente y otras más.

4.2. ANÁLISIS DEL PERFIL DE EGRESO DE LAS CARRERAS DE SOFTWARE Y TELEMÁTICA CON BASE EN LAS COMPETENCIAS ESPECÍFICAS MEDIANTE LA MODELIZACIÓN DE TÓPICOS.

El análisis del perfil de egreso de las carreras se lo realizó en dos partes, el primero un análisis descriptivo de las mallas curriculares y el segundo análisis de texto para los planes analíticos de ambas carreras.

4.2.1. Análisis descriptivo de la Malla Curricular de la carrera de Software

La malla curricular de la carrera de Software cuenta con 10 semestres revisar Anexo 2, cada uno de ellos con un mínimo de 5 asignaturas y un máximo de 7. De los 10 semestres, los tres primeros son de formación básica, y los siete segundos son de formación

profesional. En la tabla 8, se puede observar una tabla que resume las asignaturas por semestre.

Tabla 4: *Cantidad de créditos por semestre para la carrera de Software*

Semestre	Cantidad de créditos
Semestre 1	15
Semestre 2	15
Semestre 3	15
Semestre 4	14,67
Semestre 5	15
Semestre 6	15
Semestre 7	10
Semestre 8	13
Semestre 9	15
Semestre 10	15

Cada uno de los semestres cuenta con 15 créditos, a diferencia del semestre 4, 7 y 8, siendo el séptimo semestre el que cuenta con el menor número de créditos que son 10.

En cuanto a las horas y cantidad de créditos de las asignaturas de la carrera de Software se tiene que:

Tabla 5: *Estadísticos descriptivos para horas y créditos de la carrera de Software*

Estadístico	Horas	Créditos
Rango	48	2
Mínimo	48	2
Máximo	96	4
Media	70,22	2,72
Desv. Desviación	14,663	0,575

4.2.2. Análisis de la Malla curricular de la carrera de Telemática

Para el análisis de la malla curricular de la carrera de Telemática se utilizó el Anexo 3, de la cual se pudo obtener los siguientes resultados:

Tabla 6: Estadísticos descriptivos de número de horas y créditos de la carrera de Telemática

Estadístico	Hora	Crédito
Rango	16	1
Mínimo	64	2
Máximo	80	3
Media	70,55	2,5
Desviación	8,052	0,512

4.2.3. Aplicación LDA en la descripción de las competencias específicas de ambas carreras

Primero, se obtuvieron las mallas curriculares y planes analíticos de los semestres superiores a sexto de ambas carreras. Segundo, mediante la técnica de *Web Scraping* para documentos pdf se extrajeron datos como el nombre de la asignatura, cantidad de créditos, cantidad de horas, prerrequisito, descripción, objetivos, resultados de las asignaturas, temas, subtemas y metodología mismos que se almacenaron en una base de datos en Excel, revisar ilustración 14.

Ilustración 14: Extracción de la información de los planes analíticos mediante Web Scraping empleando librerías de R-studio y almacenándolos en una base de datos de Excel.



Fuente: Autor

Cada plan analítico en formato pdf corresponde a un documento diferente, los cuales mediante librería pdftool y tm de r-studio crean un solo corpus, en cada uno de ellos se aplicaron técnicas de preprocesamiento de texto, una vez limpio los datos de texto, se crea un objeto de corpus, para luego ser transformado a una Matriz de términos, a la cual se la puede realizar cualquier tipo de cálculo. Por último, estos datos son almacenados en una base de datos de Excel, la cual cuenta con variables como: Asignatura, semestre, créditos, prerequisite, descripción de la asignatura, objetivo general de la asignatura, objetivo específico de la asignatura, resultados de la asignatura.

Por medio de la modelización de tópicos se han hallado los siguientes perfiles, revisar los anexos desde el 8 al 12.

Tabla 7: *Perfiles hallados por medio de Localización Latente de Dirichlet para el perfil de egreso de los estudiantes de ambas carreras*

Perfiles	Ingenieros en Software	Ingenieros Telemáticos
	<ul style="list-style-type: none"> • Conocimiento en softwares • Programación y uso de herramientas en la nube, kubernetes • Aplicaciones móviles • Conocimiento en Linux, Android, Amazon, Azure • Analítica de datos y modelos matemáticos, frameworks, big data • Seguridad cibernética 	<ul style="list-style-type: none"> • Banda ancha • Herramientas telemáticas • Gobierno de datos • Antenas • Redes Wan, Lan • Routers • Seguridad de las redes • Proyectos y emprendimientos en telecomunicaciones

Discusión:

Empleando modelización de tópicos a los planes analíticos de ambas carreras se llegó a la conclusión que el ingeniero en software formado en la Universidad Técnica Estatal de Quevedo se gradúa con conocimientos en programación y uso de herramientas en la nube y kubernetes, aplicaciones móviles, manejo y uso de diversos sistemas operativos, análisis de datos y aplicación de modelos matemáticos y big data y con una formación básica en

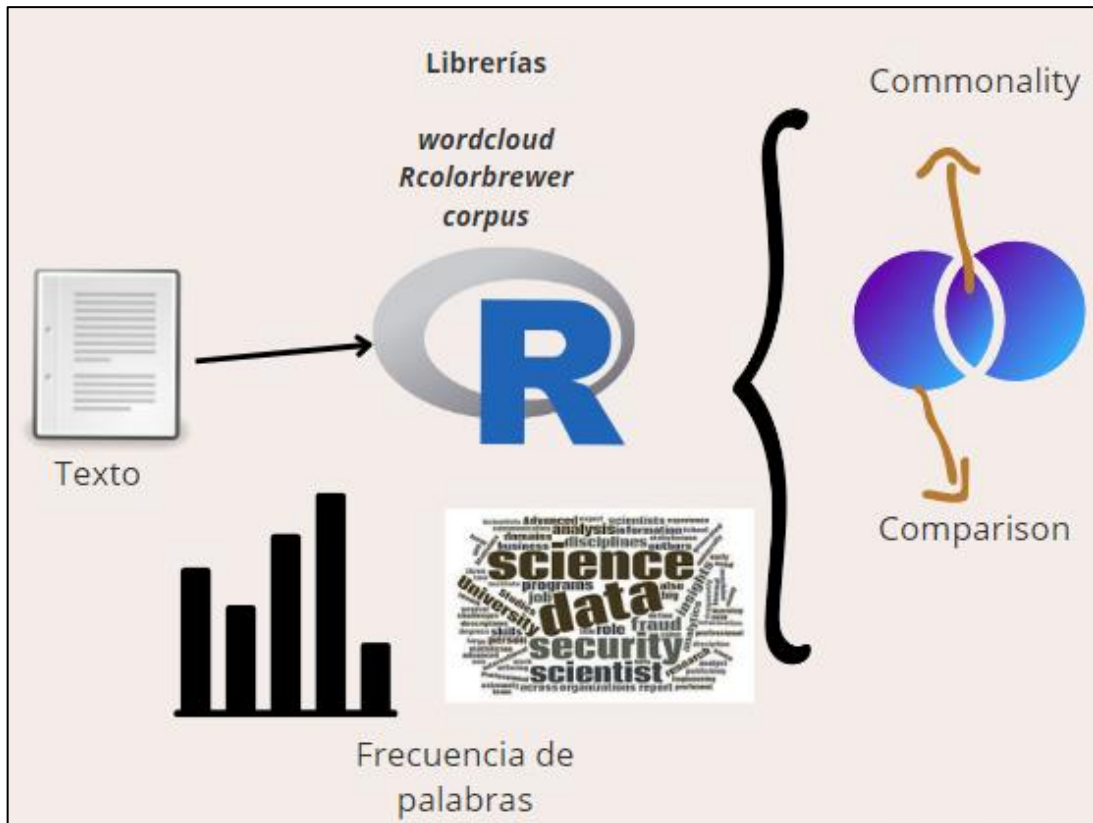
lo que respecta a la seguridad cibernética. Mientras que el ingeniero en telemática egresa con conocimientos en Banda ancha, gobierno de datos, antenas, redes WAN, LAN, Routers, Seguridad de las redes, y con bases en cuanto a la creación de proyectos y emprendimientos en lo que respecta las telecomunicaciones.

Estos perfiles tienen una formación académica, teórica y práctica que hace que los alumnos tengan un primer acercamiento a la práctica en el mundo real, de acuerdo con (Echeverría-Villafuerte & Villarruel-Meythaler, 2022) quienes en su estudio hallaron perfiles similares para las carreras relacionadas a Ciencia de datos, en donde uno de los hallazgos es la necesidad de formar profesionales con conocimientos interdisciplinarios.

4.3. ANÁLISIS DE LA BRECHA EXISTENTE ENTRE AMBOS PERFILES HALLADOS, MEDIANTE HERRAMIENTAS DE MINERÍA DE TEXTO.

Para la obtención de los resultados en esta sección, se utilizaron funciones principales del paquete Wordcloud de R-Studio el cual usa otras librerías como tm, NLP, Rcolorbrewer y otros, esta es una herramienta que crea nube de palabras que son resultado de extraer las palabras más frecuentes en un texto. *Wordcloud* trabaja con funciones entre las cuales están: *Commonality* y *Comparison*.

Ilustración 15: Proceso de la librería wordcloud junto con sus funciones de minería de texto



Fuente: Autor

Commonality es una función que entrega las palabras más frecuentes y comunes dentro de dos documentos. Mientras que la función *Comparison* compara aquellas palabras más frecuentes y distintas entre dos documentos.

A continuación, se muestran ambas gráficas en donde se puede percibir aquellas palabras comunes y diferentes entre ambos perfiles.

El perfil profesional de los estudiantes de la carrera de Software cubre con las exigencias de las vacantes publicadas en el Portal de Hiring Room en cuanto:

- Programación y uso de herramientas en la nube, kubernetes
- Aplicaciones móviles
- Conocimiento en Linux, Android, Amazon, Azure
- Analítica de datos y modelos matemáticos, frameworks, administración de base de datos
- Seguridad cibernética básica

Por otro lado, el perfil profesional de los estudiantes de la carrera de Telemática cubre con las exigencias de las vacantes publicadas en Hiring Room en cuanto a conocimientos en:

- Banda ancha
- Herramientas telemáticas
- Gobierno de datos
- Antenas
- Redes Wan, Lan
- Routers
- Seguridad de las redes
- Proyectos y emprendimientos en telecomunicaciones

Las palabras más frecuentes en el perfil de egreso de las carreras de Software y Telemática están Inteligencia Artificial, Javascript, modelos matemáticos, firebase, interfases, fromware, API Key, Linux, LAN, VLAN, switching, entre otros, términos que no han sido nombrados frecuentemente en las vacantes del Portal de Hiring Room.

Por lo tanto, existen tres brechas en el perfil de egreso de los estudiantes de Software y Telemática versus el perfil demandado por las empresas en Hiring Room que son:

- Certificaciones en ciberseguridad
- Habilidades blandas como autodesarrollo, trabajo bajo presión, atención al cliente, asertividad
- Conocimientos interdisciplinarios en otras especialidades como Finanzas, Economía, Administración.

Discusión:

Existen algunos aspectos que son factores para la presencia de brechas en cuanto al mercado laboral y los conocimientos de egresados como lo son la falta de conocimientos multidisciplinarios, habilidades blandas, y cursos complementarios que ayuden a los egresados o graduados a integrarse en el mundo laboral. De acuerdo con los resultados obtenidos por (Echeverría-Villafuerte & Villarruel-Meythaler, 2022) y (Valle Gutiérrez, 2021) es sus estudios, la brecha más grande que existe entre las exigencias del mercado laboral y la formación profesional de la universidad radica en la formación y adquisición de habilidades blandas, habilidades que muy difícilmente se incorporan en sus mallas curriculares y planes analíticos.

De acuerdo con (Varea, 2021) y (Martín Del Peso et al., 2013) quienes hacen referencia a la importancia de que las universidades formen profesionales que puedan cumplir con

las demandas exigidas por las empresas en el mercado laboral, y la constante preparación de los profesionales después de obtener su título con el fin de obtener un trabajo es cada vez más fuerte y rápida, más aún en un mundo digital y globalizado. Es por ello que indican que el profesional ideal a más de tener los conocimientos y experiencia deseada por el reclutador debe contar con habilidades blandas básicas para poder convivir con un equipo de trabajo, como ser asertivo, tener predisposición, responsabilidad, buen ambiente laboral, lealtad y trabajo en equipo, habilidades que no son impartidas en el salón de clases ni mucho menos forman parte del programa de estudios de las universidades.

Otra brecha representativa es la necesidad de un candidato con conocimientos interdisciplinarios, brecha la cual ha existido por años, y la cual ha sido raíz de la demanda de programas de maestrías, diplomados, especialidades y doctorados.

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

- En el análisis de las vacantes del portal de empleo de Hiring Room relacionadas a las carreras de Software y Telemática, se encontraron seis tipos de perfiles de profesionales requeridos. El primero, profesionales de desarrollo de Software o desarrollo de aplicaciones móviles o web. El segundo, profesionales de ciberseguridad con certificaciones o cuarto nivel. Tercero, aquellos con conocimiento en soporte técnico. Cuarto, gestores y administradores de bases de datos. Quinto, profesionales para el análisis de datos, con conocimientos de modelos de Machine Learning, estadística, matemática, economía, finanzas. Sexto profesionales de telemática y telecomunicaciones con conocimientos en redes, cableado estructurado, soporte técnico. En cuanto a las empresas que más requieren de estos profesionales, está el sector financiero, seguido del sector tecnológico, mismas que se encuentran ubicadas en las principales ciudades del país como Quito, Guayaquil y Cuenca.
- En cuanto a los perfiles de egreso de los estudiantes de Software y Telemática de la Universidad Técnica Estatal de Quevedo se forman con buenas bases de conocimiento en: Programación y uso de herramientas de la nube, aplicaciones móviles, sistemas operativos, analítica de datos, frameworks, seguridad cibernética de forma básica en el caso de Software, para Telemática, los estudiantes obtienen sólidos conocimientos en Banda Ancha, Gobernabilidad de datos, antenas, redes WAN, LAN, routers, seguridad de las redes, proyectos y emprendimiento en telecomunicaciones
- Se encontraron 3 brechas en el perfil de egreso de las carreras de Software y Telemática versus las exigencias del mercado laboral, las cuales son:

Certificaciones en ciberseguridad, habilidades blandas y conocimientos interdisciplinarios.

5.2. Recomendaciones

- Se recomienda aplicar la metodología empleada a otros portales de empleo o con un enfoque diferente cómo por ejemplo analizar el perfil demandado por las empresas para empleos remotos con el fin de descubrir patrones que brinden una guía a los lectores sobre los requerimientos para este tipo de empleos.
- Se recomienda que la Universidad Técnica Estatal de Quevedo cree programas de integración o acompañamientos psicológicos para poder adquirir aquellas habilidades blandas para que sus graduados puedan entrar al mercado laboral sin necesidad de tener estudios complementarios o de cuarto nivel, de igual forma crear capacitaciones que incluyan la entrega de certificaciones no solo en el área de ciberseguridad como lo mostraron los resultados, sino también en otros aspectos, para que en el momento que sus estudiantes se gradúen ya cuenten con estos, dándoles ventaja competitiva.
- Se recomienda a la Universidad Técnica Estatal de Quevedo crear programas o talleres de vinculación con la sociedad donde los estudiantes de la Facultad de Ciencias de la Ingeniería se relacionen con otros estudiantes de otras facultades con el fin de reducir la brecha en cuando a conocimientos multidisciplinarios.

REFERENCIAS BIBLIOGRÁFICAS

- AnalyStats. (2019). *Análisis de datos: Latent Dirichlet Allocation (LDA) Aplicada en Textos Periodísticos*. <https://blog.analystats.com/2019/05/02/an%C3%A1lisis-de-datos-latent-dirichlet-allocation-lda-aplicada-en-textos-period%C3%ADsticos/>
- Annisa, R., & Surjandari, I. (2019). Opinion Mining on Mandalika Hotel Reviews Using Latent Dirichlet Allocation. *Procedia Computer Science*, 161, 739–746. <https://doi.org/10.1016/j.procs.2019.11.178>
- AVI Networks. (s. f.). *What is Web Scraping?* Recuperado 30 de septiembre de 2023, de <https://avinetworks.com/glossary/web-scraping/>
- Bechmann, A., & Bowker, G. C. (2019). Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. *Big Data and Society*, 6(1). <https://doi.org/10.1177/2053951718819569>
- Bernal-Yamuca, J. L., Fuentes-Indio, C. M., Sion-Ganchozo, N. Y., & Tapia-Miranda, E. E. (2022). Análisis de los principales indicadores del mercado laboral y las ramas de ocupación en el Ecuador. *Ciencia Latina Revista Científica Multidisciplinar*, 6(5), 903–924. https://doi.org/10.37811/cl_rcm.v6i5.3151
- Blei, D. M., Ng, A. Y., & Edu, J. B. (2003). Latent Dirichlet Allocation Michael I. Jordan. *Journal of Machine Learning Research*, 3, 993–1022.
- Buenano-Fernandez, D., Gonzalez, M., Gil, D., & Lujan-Mora, S. (2020). Text Mining of Open-Ended Questions in Self-Assessment of University Teachers: An LDA Topic Modeling Approach. *IEEE Access*, 8, 35318–35330. <https://doi.org/10.1109/ACCESS.2020.2974983>
- Castillo, A., & Aguilar, G. (2020). *DESARROLLO DE SISTEMA DE ANÁLISIS DE EMPLEABILIDAD EN PORTALES WEB DE EMPLEOS*.
- Chauhan, U., & Shah, A. (2021). Topic Modeling Using Latent Dirichlet allocation. *ACM Computing Surveys*, 54(7). <https://doi.org/10.1145/3462478>
- Contreras - Barrera, M. (2014). *Minería de texto: una visión actual* (Vol. 17, Número 2). <http://www.emc.com/leadership/programs/digital-universe.htm>

- Diaz-Rubiano, M. A. (2022). ANÁLISIS DE TEMAS UTILIZANDO TWITTER: UNA APLICACIÓN DEL MODELO LDA AL CASO COLOMBIANO. *Universidad Santo Tomás*, 1–80.
- Dogucu, M., & Çetinkaya-Rundel, M. (2020). Web Scraping in the Statistics and Data Science Curriculum: Challenges and Opportunities. *Journal of Statistics Education*. <https://doi.org/10.1080/10691898.2020.1787116>
- Dympna Kelly, Siperstein Allan, & Fung John J. (2015). *A Structured Educational Curriculum Including Online Training Positively Impacts American Board of Surgery In-Training Examination Scores*. Elsevier. <https://www.sciencedirect.com/science/article/abs/pii/S1931720415001038>
- Echeverría-Villafuerte, D. H., & Villarruel-Meythaler, R. Efraín. (2022). *Modelización de tópicos con localización latente de Dirich-let en la caracterización entre el perfil laboral y de egreso del Programa de Estadística de la Universidad Central del Ecuador durante el periodo octubre 2020-julio 2021*.
- Figure from Latent Dirichlet Allocation (LDA) for Topic Modeling of the CFPB Consumer Complaints Semantic Scholar*. (2023). https://sitelikeet.life/product_details/15883979.html
- Ilyés, V., & Seb, A. (2023). *University peers and career prospects: The impact of university ties on early labor market outcomes*. <https://doi.org/10.1016/j.econedurev.2023.102456>
- Info Capital Humano. (s. f.). *Hiring Room: software reduce el proceso de selección y contratación de personal*. Recuperado 7 de octubre de 2023, de <https://infocapitalhumano.pe/recursos-humanos/noticias-y-movidas/hiring-room-software-reduce-el-proceso-de-seleccion-y-contratacion-de-personal/>
- Islam, D., Li, B., Islam, K. S., Ahasan, R., Mia, R., & Haque, E. (2022). Airbnb rental price modeling based on Latent Dirichlet Allocation and MESF-XGBoost composite model. *Elsevier*, 7, 1–9. <https://doi.org/10.1016/j.mlwa.2021.100208>
- Laboratorio de Datos. (2021). *Laboratorio de Datos: Análisis de tópicos (LSA, NMF, LDA)*. Youtube. <https://www.youtube.com/watch?v=nNwMIOo4q-U>
- Laia, M., & Calvo, M. (2019). *Web scraping*.

- Ley de comercio electrónico, firmas y mensajes de datos (2002). www.lexis.com.ec
- Ley Orgánica de Protección de Datos Personales (2021). www.lexis.com.ec
- Ley Orgánica de Telecomunicaciones (2015).
- Lupi, F., & Mabkhot, M. (2023). *Automatic definition of engineer archetypes: A text mining approach / Enhanced Reader*. Elsevier. <https://www.sciencedirect.com/science/article/pii/S016636152300146X>
- Luque, C., Rubriche, J., Galvis, J., & Sosa, J. (2021). Topic modeling to identify patterns in Covid-19 scientific research. En *Comunicaciones en Estadística Diciembre* (Vol. 14, Número 2).
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures*, 12(2–3), 93–118. <https://doi.org/10.1080/19312458.2018.1430754>
- Malla Curricular Software - Universidad Tecnica Estatal de Quevedo*. (2024). <https://www.uteq.edu.ec/es/grado/carrera/software>
- Malla Curricular Telematica - Universidad Tecnica Estatal de Quevedo*. (2024). <https://www.uteq.edu.ec/assets/images/careers/degrees/academic/img-uteq-malla-000029.jpg>
- Martín Del Peso, M., Rabadán Gómez, A. B., & Hernández March, J. (2013). Desajustes entre formación y empleo en el ámbito de las enseñanzas técnicas universitarias: La visión de los empleadores de la Comunidad de Madrid. *Revista de Educacion*, 360, 244–267. <https://doi.org/10.4438/1988-592X-RE-2011-360-110>
- Paterson, L. (2023). Breadth of study at secondary school and the attainment and progression of university graduates in Scotland, 1960–2002. *Higher Education*, 85(4), 717–738. <https://doi.org/10.1007/S10734-022-00862-4>
- Pietsch, A. S., & Lessmann, S. (2018). Topic modeling for analyzing open-ended survey responses. *Journal of Business Analytics*, 1(2), 93–116. <https://doi.org/10.1080/2573234X.2019.1590131>

- Prabha, S., & Sardana, N. (2023). Question Tags or Text for Topic Modeling: Which is better. *Procedia Computer Science*, 218, 2172–2180. <https://doi.org/10.1016/J.PROCS.2023.01.193>
- Ramirez, J. (2022). Análisis Automatizado de políticas de privacidad en el Ecuador. En *Escuela Politecnica Nacional*.
- Ridge, B. (2024). *Factores clave que los empleadores buscan en un profesional de diseño de multimedia*. <https://www.mediummultimedia.com/disenio/que-buscan-los-empleadores-en-un-disenador-de-multimedia/>
- Ruiz Díaz de Salvioni, V. V., & Armoa, A. (2023). La importancia de la Minería de Datos como una herramienta estratégica en las Empresas. *Ciencia Latina Revista Científica Multidisciplinar*, 7(1), 9267–9276. https://doi.org/10.37811/cl_rcm.v7i1.5119
- Samaniego, E. (2023). *WEB MINING*.
- Sans, L., Vallvé, I., Teixidó, J., Picas, J. M., Martínez-Roldán, J., & Pascual, J. (2022). La era del big data: análisis del lenguaje natural mediante la aplicación de folksonomía. *Sociedad Española de Nefrología*, 42(6), 680–687. <https://doi.org/10.1016/J.NEFRO.2021.09.006>
- Seah, K. K. C., Pan, J., & Tan, P. L. (2020). Breadth of university curriculum and labor market outcomes. *Labour Economics*, 65, 101873. <https://doi.org/10.1016/J.LABECO.2020.101873>
- Sierra, V. (2022). El seguimiento a graduados: un vínculo entre las instituciones de educación superior y la sociedad. *Scielo*, 1–17. <https://orcid.org/0000-0002-0465-2212>
- Slof, D., Frasincar, F., & Matsiako, V. (2021). A competing risks model based on latent Dirichlet Allocation for predicting churn reasons. *Elsevier*. <https://doi.org/10.1016/j.dss.2021.113541>
- Solé-Moro, M. L., Sánchez-Torres, J. A., Arroyo-Cañada, F. J., & Argila-Irurita, A. (2018). Los egresados universitarios y la inserción laboral: un acercamiento al panorama latinoamericano y español. *Revista CEA*, 4(8), 67–74. <https://doi.org/10.22430/24223182.1048>

- Stenhouse, N. V. (2018). *HabScraper: herramienta automatizada para la extracción de datos con web scraping*.
- Terreros-Ríos, A. J., Vega-Villavicencio, A. J., & Pupo-Francisco, J. M. (2019). Aplicación del Teorema de Bayes en la selección de personal para disminuir la deserción laboral. *593 Digital Publisher CEIT*, 6(4), 27–40. <https://doi.org/10.33386/593dp.2019.6.140>
- Valle Gutiérrez, G. (2021). *APLICACIÓN DE TÉCNICAS DE WEB SCRAPING Y PROCESAMIENTO DEL LENGUAJE NATURAL PARA LA EXTRACCIÓN Y EVALUACIÓN DE INFORMACIÓN DE UNA PÁGINA WEB DE EMPLEO*.
- Varea, R. (2021). *De la universidad a la empresa: así se facilita a los estudiantes su acceso al empleo* . <https://elpais.com/sociedad/educacion-online/2021-10-25/de-la-universidad-a-la-empresa-asi-se-facilita-a-los-estudiantes-su-acceso-al-empleo.html>
- Vernazza, E. (2020). *¿Qué se publica sobre el Coronavirus? Una propuesta de modelado de tópicos mediante LDA*. Semana Internacional de la Estadística. <https://www.youtube.com/watch?v=nU7kkvQfzRE>
- Wang, L. (2022). The gap between university studies and labour market in journalism: opening educational community example in China. *CelPress*. <https://doi.org/10.1016/j.heliyon.2022.e11973>

Anexos

Anexo 1: Informe Compilatio – Antiplagio

PARA: Dr. Byron Oviedo Bayas
Decano de Posgrado
DE: Ing. Emilio Zhuma Mera, Ms.C
ASUNTO: Informe Proyecto de Investigación
FECHA: 1 de abril de 2024

Adjunto al presente sírvase encontrar el documento final del proyecto de investigación titulado: **CARACTERIZACIÓN DEL PERFIL DE EGRESO DE LOS PROFESIONALES DE LA UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO**, elaborado por la **ING. MARIA ISABEL CHEVEZ CASTRO** posgradista de la **MAESTRÍA EN CIENCIA DE DATOS**. El proyecto de investigación fue elaborado bajo mi dirección según lo asignado en el contrato Nro. UTEQ-RUTEQ- 2023- 4186-M de fecha 24 de noviembre de 2023, el mismo que cumple el informe de la herramienta COMPILATIO, el cual avala los niveles de originalidad, en un 98 % del trabajo investigativo.

 CERTIFICADO DE ANÁLISIS
magister

Prueba antiplagio - Maria Chevez

2%
Textos sospechosos

< 1% Similitudes
0% similitudes entre comillas
0% entre las fuentes mencionadas

2% Idiomas no reconocidos

Nombre del documento: Prueba antiplagio - Maria Chevez.pdf	Depositante: EMILIO RODRIGO ZHUMA MERA	Número de palabras: 13.162
ID del documento: eee240fdd22fbc4a291667290a3ed4a154a968ef	Fecha de depósito: 1/4/2024	Número de caracteres: 86.288
Tamaño del documento original: 680,7 kB	Tipo de carga: interface	
	fecha de fin de análisis: 1/4/2024	

Ubicación de las similitudes en el documento:





Atentamente,

EMILIO
RODRIGO
ZHUMA
MERA

Firmado digitalmente por EMILIO RODRIGO ZHUMA MERA

Ing. Emilio Zhuma Mera, Ms.C.
Director de Tesis

Anexo 2: Malla curricular de Software

 MALLA CURRICULAR SOFTWARE 									
I Semestre	II Semestre	III Semestre	IV Semestre	V Semestre	VI Semestre	VII Semestre	VIII Semestre	IX Semestre	X Semestre
ÁLGEBRA LINEAL 3	ARQUITECTURA DE COMPUTADORAS 2	BASE DE DATOS 2	DINÁMICA DE GRUPO Y COMUNICACIÓN 1.07	ADMINISTRACIÓN DE BASES DE DATOS 3	ADMINISTRACIÓN DE PLATAFORMAS DE SOFTWARE 3	APLICACIONES DISTRIBUIDAS 2	DEONTOLOGÍA INFORMÁTICA 3	GENERALES ECONÓMICA, SOCIAL Y E INNOVACIÓN 3	DESARROLLO DEL TRABAJO DE TITULACIÓN I 4
ALGORITMOS Y LÓGICA DE PROGRAMACIÓN 3	CÁLCULO INTEGRAL 2	ESTRUCTURA DE DATOS 3	FUNDAMENTOS DE EXINABILIDAD Y FINANZAS 2	APLICACIONES WEB 3	APLICACIONES MÓVILES 3	CALIDAD DE SOFTWARE 2	GESTIÓN DE LA CONFIGURACIÓN DE SOFTWARE 2	FORMULACIÓN DEL TRABAJO DE TITULACIÓN 4	LEGISLACIÓN INFORMÁTICA 3
CÁLCULO DIFERENCIAL 3	ESTRUCTURAS DISCRETAS II 3	MÉTODOS NUMÉRICOS 2	INGENIERÍA DE REQUERIMIENTOS 3	DISEÑO Y ARQUITECTURA DE SOFTWARE 3	CONSTRUCCIÓN DE SOFTWARE 3	INTELIGENCIA DE NEGOCIOS 2	HERRAMIENTAS DE PROGRAMACIÓN 3	GESTIÓN DE TECNOLOGÍAS DE LA INFORMACIÓN 3	LIDERAZGO Y EMPRENDIMIENTO 2
ENFOQUE A LA INGENIERÍA DE SOFTWARE 3	FUNDAMENTOS FÍSICA PARA LA INGENIERÍA 2	MODELAMIENTO DE SOFTWARE 3	INTERACCION HOMBRE MAQUINA 3	MÓDULOS MATEMÁTICOS Y SIMULACIÓN 3	PROCESO DE SOFTWARE 3	MANTENIMIENTO DE SOFTWARE 2	INGENIERÍA ECONÓMICA PARA SOFTWARE 2	REALIDAD SOCIOECONÓMICA GLOBAL DEL DELEGADO DEL MUNDO 3	PROGRAMACIÓN EN LA NUBE 3
ESTRUCTURAS DISCRETAS I 3	METODOLOGÍA DE LA INVESTIGACIÓN 3	PROBABILIDAD Y ESTADÍSTICA 2	INVESTIGACIÓN DE OPERACIONES 3	VERIFICACIÓN Y VALIDACIÓN DE SOFTWARE 3	REDES DE DATOS 3	SEGURIDAD INFORMÁTICA 2	MÉTODOS DE ELEGIR Y OPTIMIZACIÓN 3	SISTEMAS INTELIGENTES 2	REDACCIÓN DE DOCUMENTOS CIENTÍFICOS 3
INGLÉS I	INGLÉS II	INGLÉS III	INGLÉS IV	INGLÉS V	INGLÉS VI 0	INGLÉS VI	INGLÉS VII		

Importante

Formación Básica ■

Formación Profesional ■

Formación Integración Curricular ■

ASIGNATURA 0



CRÉDITOS

Activar Windows

Ve a Configuración para activar Windows.

Fuente: (Malla Curricular Software - Universidad Técnica Estatal de Quevedo, 2024)

Anexo 3: Malla Curricular de la Carrera de Telemática

 MALLA CURRICULAR TELEMÁTICA 									
I Semestre	II Semestre	III Semestre	IV Semestre	V Semestre	VI Semestre	VII Semestre	VIII Semestre	IX Semestre	X Semestre
ÁLGEBRA LINEAL 2	CÁLCULO INTEGRAL 3	CÁLCULO VECTORIAL 3	DISEÑO DE BASE DE DATOS 3	ANÁLISIS DE SEÑALES Y SISTEMAS 3	MÉTODOS ESTADÍSTICOS PARA LAS TELECOMUNICACIONES 3	ADMINISTRACIÓN Y GESTIÓN DE SERVICIOS Y SERVICIOS 2	APLICACIONES TELEMÁTICAS 2	COMUNICACIONES MÓVILES 3	ANTENAS Y PROPAGACIÓN 3
ALGORITMOS 3	CIRCUITOS ELÉCTRICOS 3	ECUACIONES DIFERENCIALES 3	ELECTRÓNICA DIGITAL 3	ÉTICA PROFESIONAL 2	PROCESAMIENTO DIGITAL DE SEÑALES 3	INVESTIGACIÓN DE OPERACIONES 2	APLICACIONES TELEMÁTICAS BASADAS EN WEB 2	DESARROLLO DE TITULACIÓN I 4	DESARROLLO DE TITULACIÓN I 4
CÁLCULO DIFERENCIAL 3	FÍSICA APLICADA 3	ELECTRICIDAD Y MAGNETISMO 3	MATEMÁTICA APLICADA 3	PROBABILIDADES Y ESTADÍSTICA 2	REDES ORDENADORES 3	ROBÓTICA 2	POLÍTICA CIENTÍFICA TECNOLÓGICA 2	ECONOMÍA DEL CAMBIO TECNOLÓGICO 2	ENTENDIMIENTOS E INNOVACIONES TECNOLÓGICAS 3
FÍSICA BÁSICA 3	FUNDAMENTOS DE PROGRAMACIÓN 3	ELECTRÓNICA ANALÓGICA 3	MÉTODOS NUMÉRICOS 3	PROGRAMACIÓN DE BASE DE DATOS 3	SISTEMAS DE COMUNICACIONES DIGITALES 3	SISTEMAS OPERATIVOS DE RED 2	REDES ESCALABLES 2	INTEGRACIÓN DE REDES 3	PROGRAMACIÓN DE PROGRAMACIÓN EN APLICACIONES MATEMÁTICAS 2
LÓGICA MATEMÁTICA 2	METODOLOGÍA DE INVESTIGACIÓN 3	PROGRAMACIÓN ORIENTADA A OBJETOS 3	TEORÍA ELECTROMAGNÉTICA 3	SISTEMAS DE COMUNICACIONES ANALÓGICAS 3	SISTEMAS EMBEBIDOS 3	TELEMÁTICA 2	SEGURIDAD EN REDES Y SERVICIOS 2	SISTEMAS MÓVILES Y OBJETOS 3	PROYECTOS DE TELECOMUNICACIONES 3
REDACCIÓN TÉCNICA 2				SISTEMAS OPERATIVOS 2			SEÑALIZACIÓN Y CONMUTACIÓN 2		
INGLÉS I	INGLÉS II	INGLÉS II	INGLÉS IV	INGLÉS V	INGLÉS VI	INGLÉS VII	INGLÉS VIII		

Importante

Formación Básica ■

Formación Profesional ■

Formación Integración Curricular ■

ASIGNATURA 0

CRÉDITOS

Fuente: (Malla Curricular Telemática - Universidad Técnica Estatal de Quevedo, 2024)

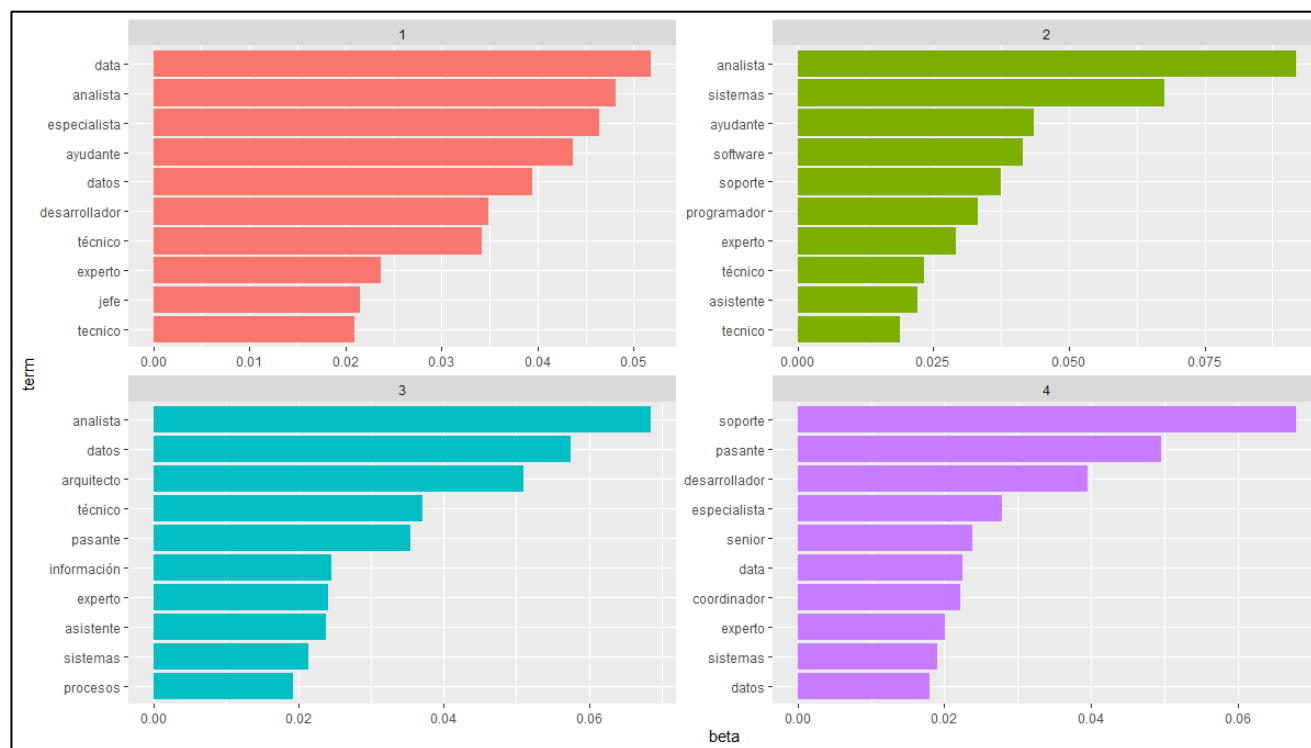
Anexo 4: Ejemplo de la obtención de las probabilidades Beta y gammas de que una palabra clave pertenezca a alguno de los tópicos establecidos en los objetivos de cada asignatura y las probabilidades de que cada documento pertenezca a un tópico

	topic	term	beta
1	5	software	0.066844920
2	6	software	0.061776062
3	2	aplicaciones	0.050465051
4	2	información	0.044040271
5	2	principales	0.037442786
6	4	asignatura	0.031275592
7	2	datos	0.031185789
8	6	desarrollo	0.030888031
9	2	plataforma	0.023827227
10	6	herramientas	0.023166023
11	6	construcción	0.023166023
12	6	redes	0.023166023
13	4	red	0.022247180
14	2	desarrollo	0.022034934
15	5	asignatura	0.021390374
16	5	calidad	0.021390374
17	2	tecnologías	0.021315507
18	1	señales	0.020491803
19	2	comunicación	0.020423338
20	4	sistemas	0.019775271
21	6	ciclo	0.019305019
22	6	sistemas	0.019305019

Showing 1 to 22 of 5,688 entries, 3 total columns

	document	topic	gamma
1	1	1	0.0002356323
41	1	2	0.0002356323
81	1	3	0.0002356323
121	1	4	0.0002356323
161	1	5	0.9988218387
201	1	6	0.0002356323
10	10	1	0.0001890766
50	10	2	0.0001890766
90	10	3	0.0001890766
130	10	4	0.0001890766
170	10	5	0.0001890766
210	10	6	0.9990546172
11	11	1	0.0004907084
51	11	2	0.0004907084
91	11	3	0.9975464582
131	11	4	0.0004907084

Anexo 5: Creación de tópicos con el Modelo de Localización Latente de Dirichlet a las descripciones de las vacantes

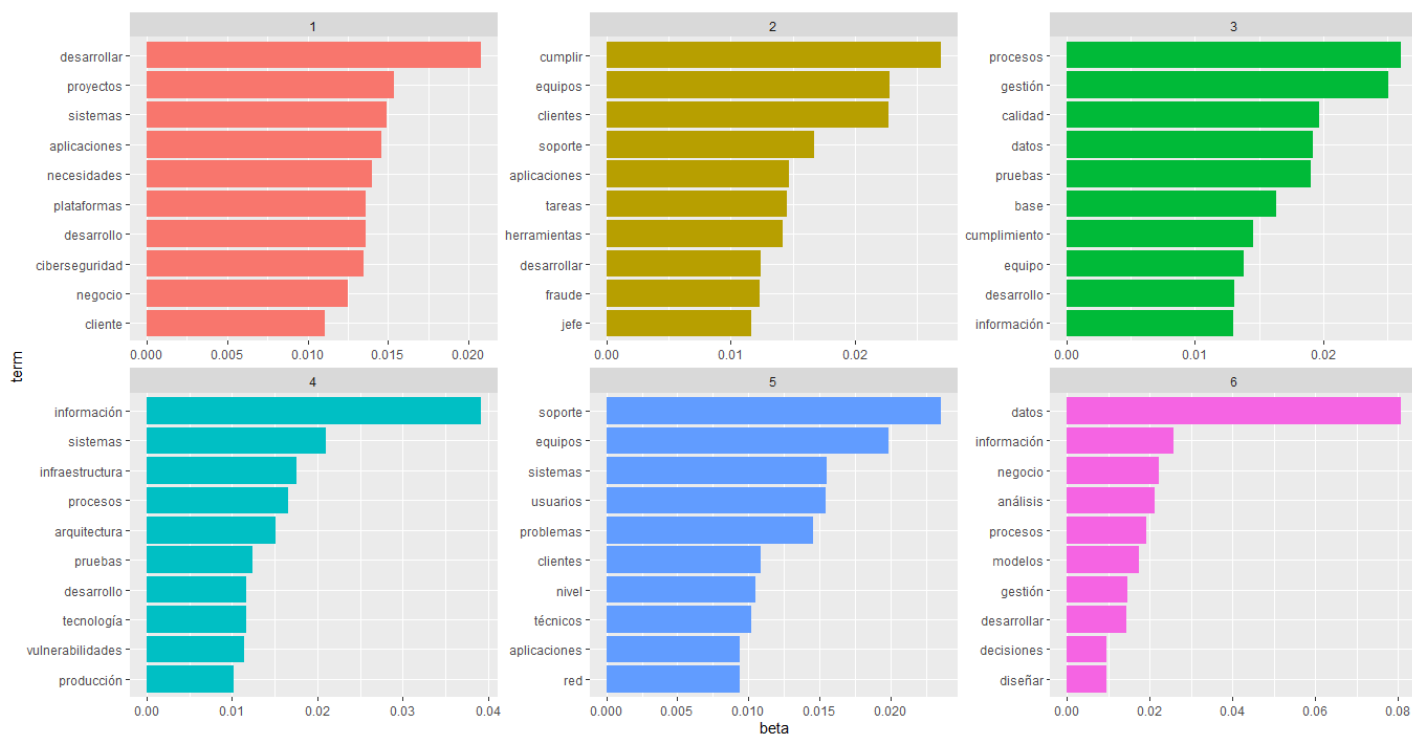


Tópico 1: Vacantes relacionadas a la analítica de datos, jefes técnicos ayudantes, especialistas

Tópico 2: Analistas de sistemas y desarrolladores de software, asistente, ayudante, técnico

Tópico 3: Analista y arquitectos de data, incluyen asistentes expertos y pasantes

Tópico 4: Soporte técnico, desarrolladores, expertos, senior y pasantes



Tópico 1: Desarrollar sistemas y aplicaciones, proyectos de negocio relacionados a la ciberseguridad

Tópico 2: Dar soporte al equipo y al cliente

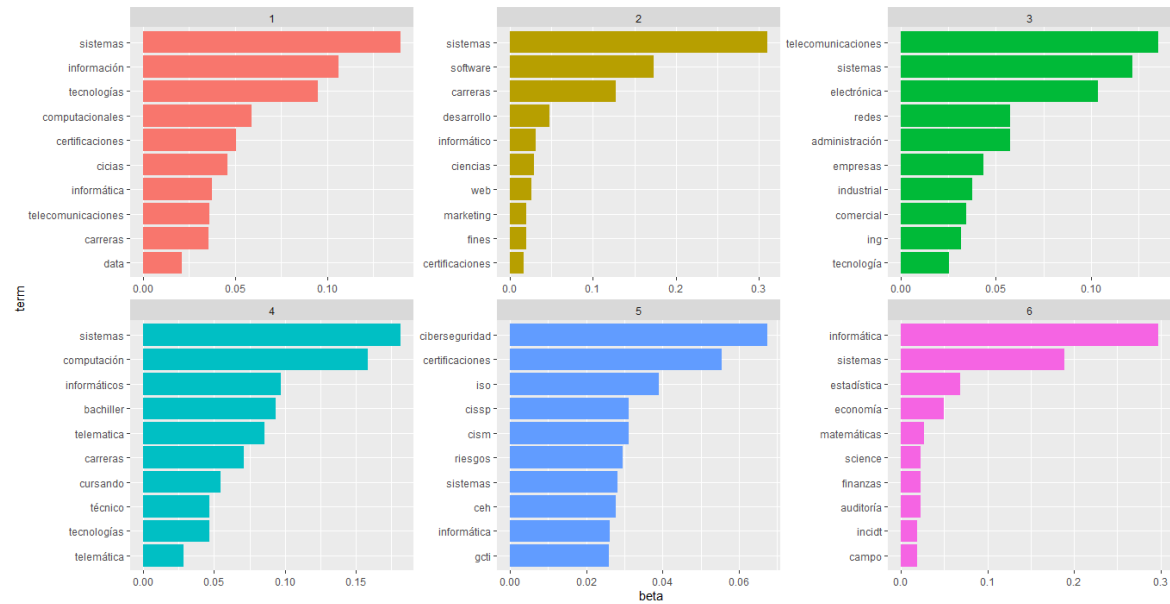
Tópico 3: Gestionar y administrar las bases de datos

Tópico 4: Dar soporte en cuanto a la infraestructura y la arquitectura de los negocios, reconociendo las vulnerabilidades existentes

Tópico 5: Dar soporte técnico a clientes y usuarios de la red

Tópico 6: Analizar los datos y crear modelos que ayuden a la toma de decisiones

Anexo 6: Creación de tópicos por medio de Localización Latente de Dirichlet para los requisitos de educación



Tópico 1: Tecnologías computacionales, telecomunicaciones certificaciones

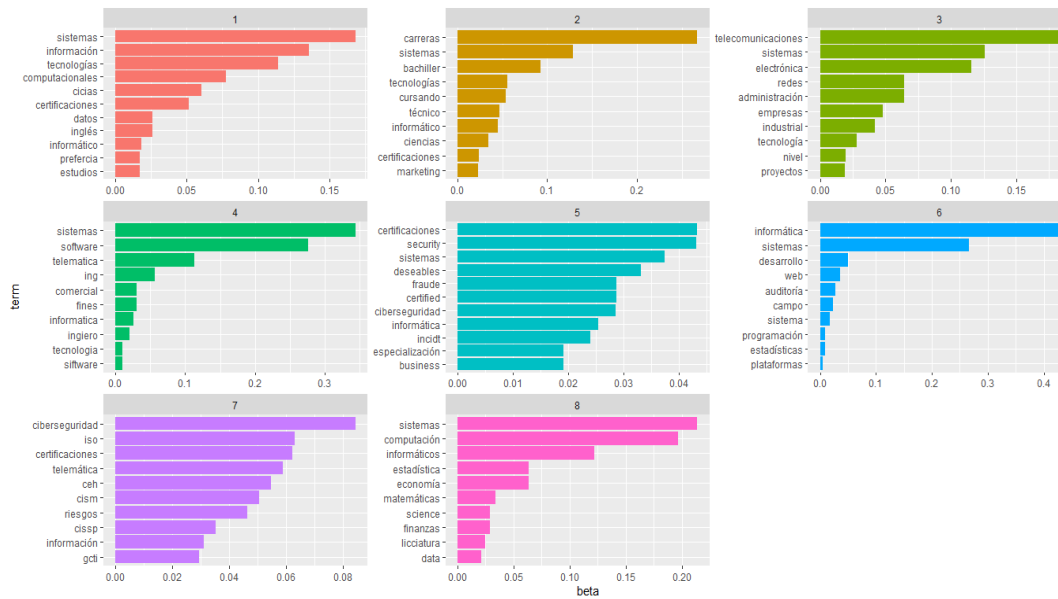
Tópico 2: Certificaciones, desarrollos informáticos, marketing web

Tópico 3: Telecomunicaciones, electrónica, redes, administración, de empresas industrias

Tópico 4: Sistemas informáticas, telemática, bachilleres o cursando la carrera

Tópico 5: Ciberseguridad certificaciones: ISO, CISSP, CISM, CEH, GCTIm sistemas informáticos

Tópico 6: Data science, estadística, modelos matemáticos y económicos, finanzas básicas



Tópico 1: Tecnologías computacionales, certificaciones, idioma inglés, manejo de datos

Tópico 2: Cursando carreras tecnológicas o bachilleres, conocimiento en informática y marketing y con certificaciones

Tópico 3: Telecomunicaciones, redes y electrónica, proyectos industriales

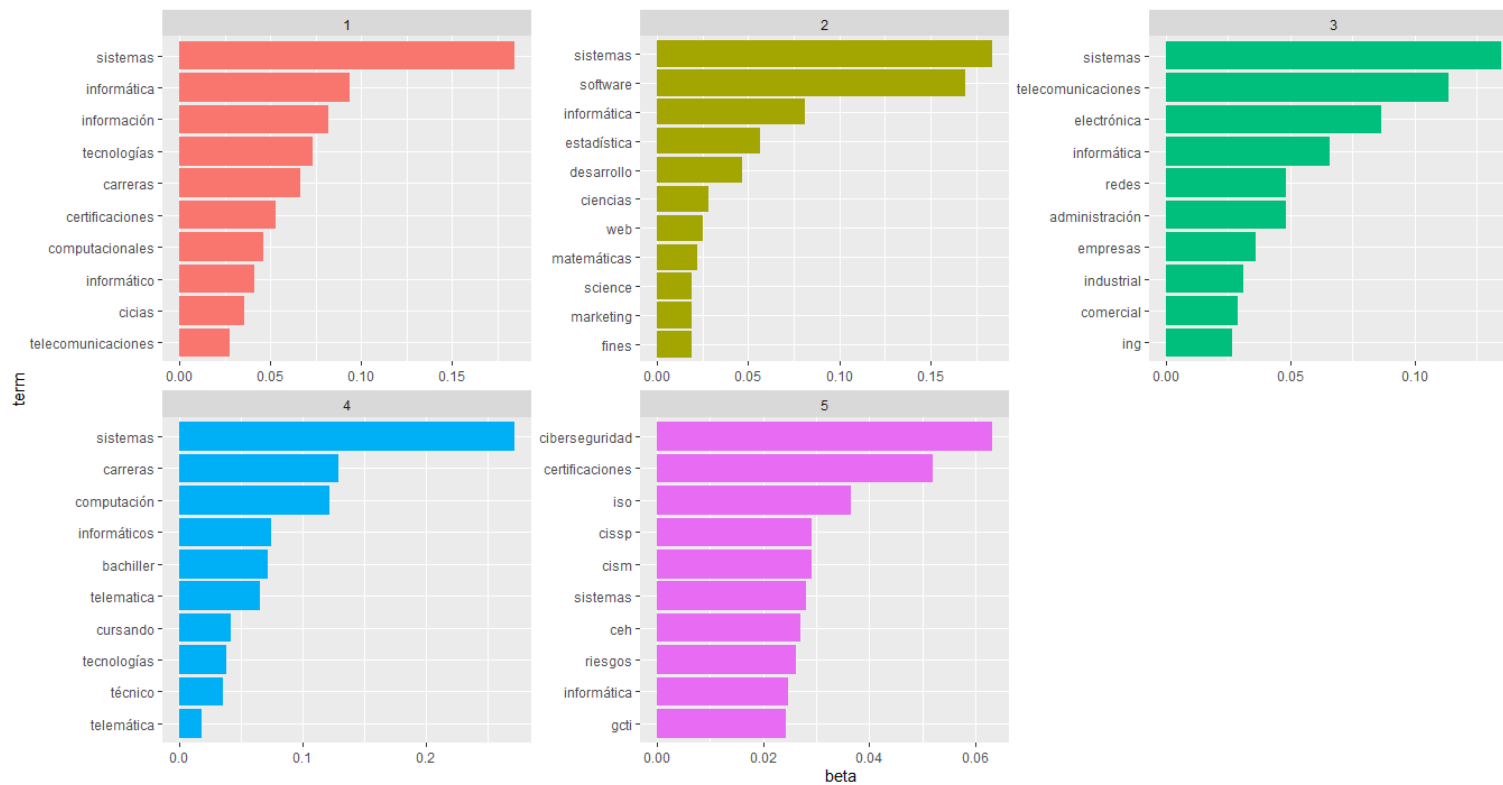
Tópico 4: Sistemas, software y telemática

Tópico 5: Ciberseguridad, certificaciones o especializaciones, fraude, incidentes

Tópico 6: Desarrollo sistemas web, programación en el campo de la estadística, auditoria web

Tópico 7: Telemáticos especializados en ciberseguridad certificaciones como ISO, CEH, CISM, CISSP, GCTI

Tópico 8: Data science con conocimiento en estadísticas, modelos matemáticos y económicos, programación



Tópico 1: Telecomunicaciones, informática, datos,

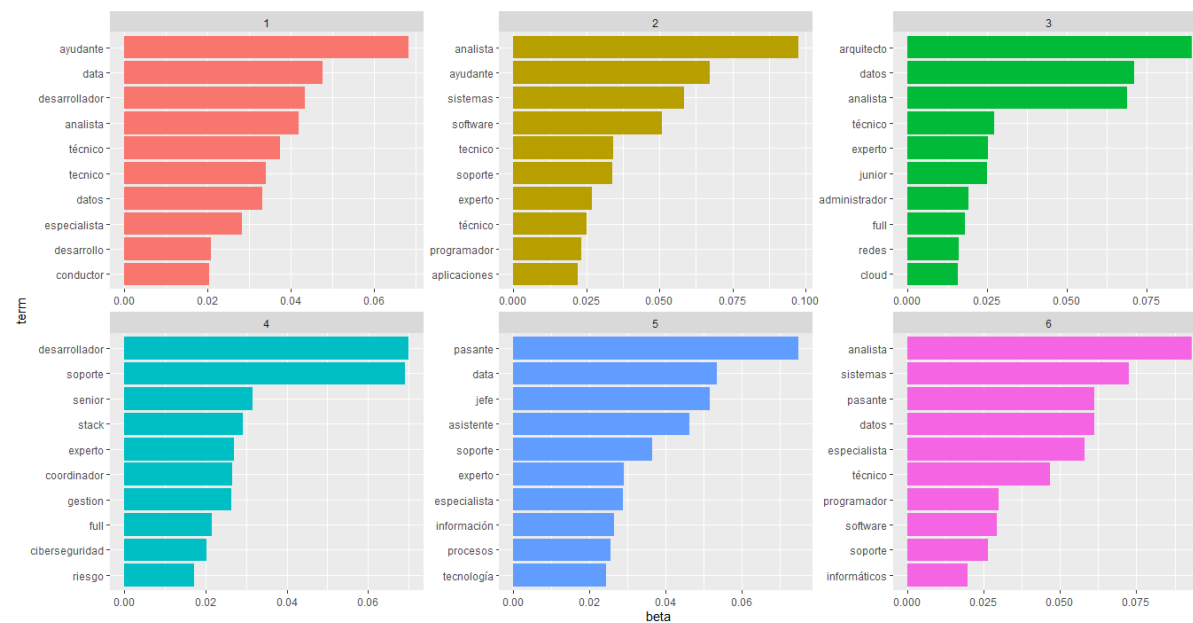
Tópico 2: Data science, estadística, matemática, marketing digital,

Tópico 3: Telecomunicaciones, informática, electrónica, administración de empresas, industrias

Tópico 4: Pasantes telemática, o cursando carreras tecnológicas

Tópico 5: Ciberseguridad con certificaciones ISO, CEH, CISM, CISSP, GCTI

Anexo 7: Creación de tópicos mediante Localización Latente de Dirichlet a los requisitos de conocimiento



Tópico 1: Desarrollo y conocimientos en datos, conductor

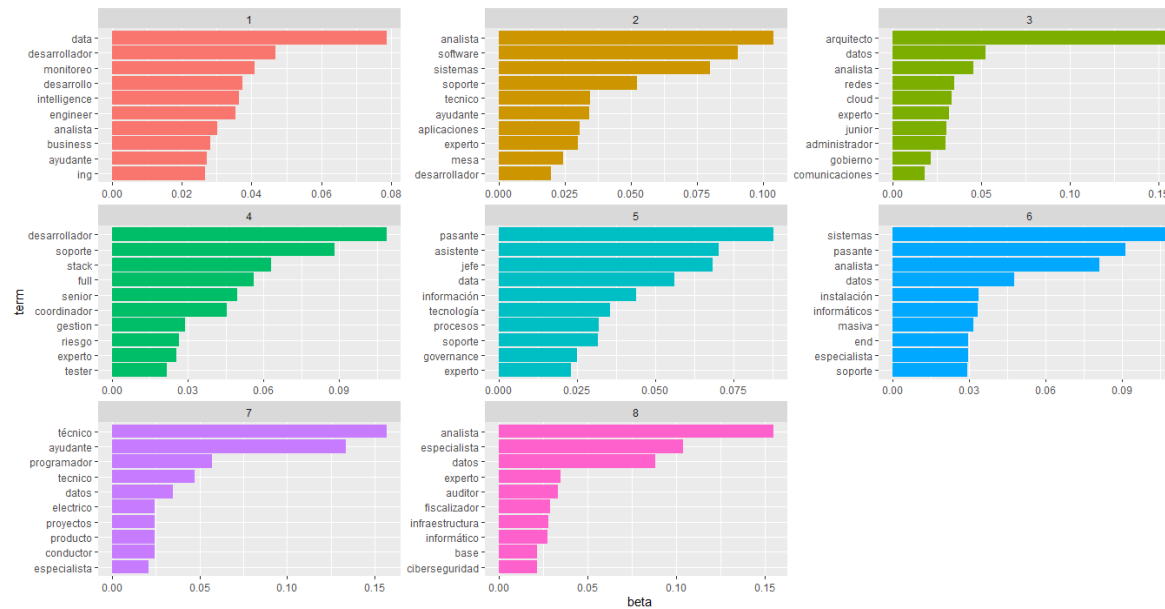
Tópico 2: Conocimientos en soporte técnico programación y desarrollo de software

Tópico 3: Arquitectura de datos, y redes, conocimiento en herramientas de la nube

Tópico 4: Ciberseguridad

Tópico 5: Datos, soporte técnico, información, procesos y tecnología

Tópico 6: Datos y programación



Tópico 1: Business intelligence, monitoreo y análisis de datos

Tópico 2: Análisis de calidad de software, mesa y soporte técnicos

Tópico 3: Arquitectura de datos, herramientas en la nube, gobierno de datos,

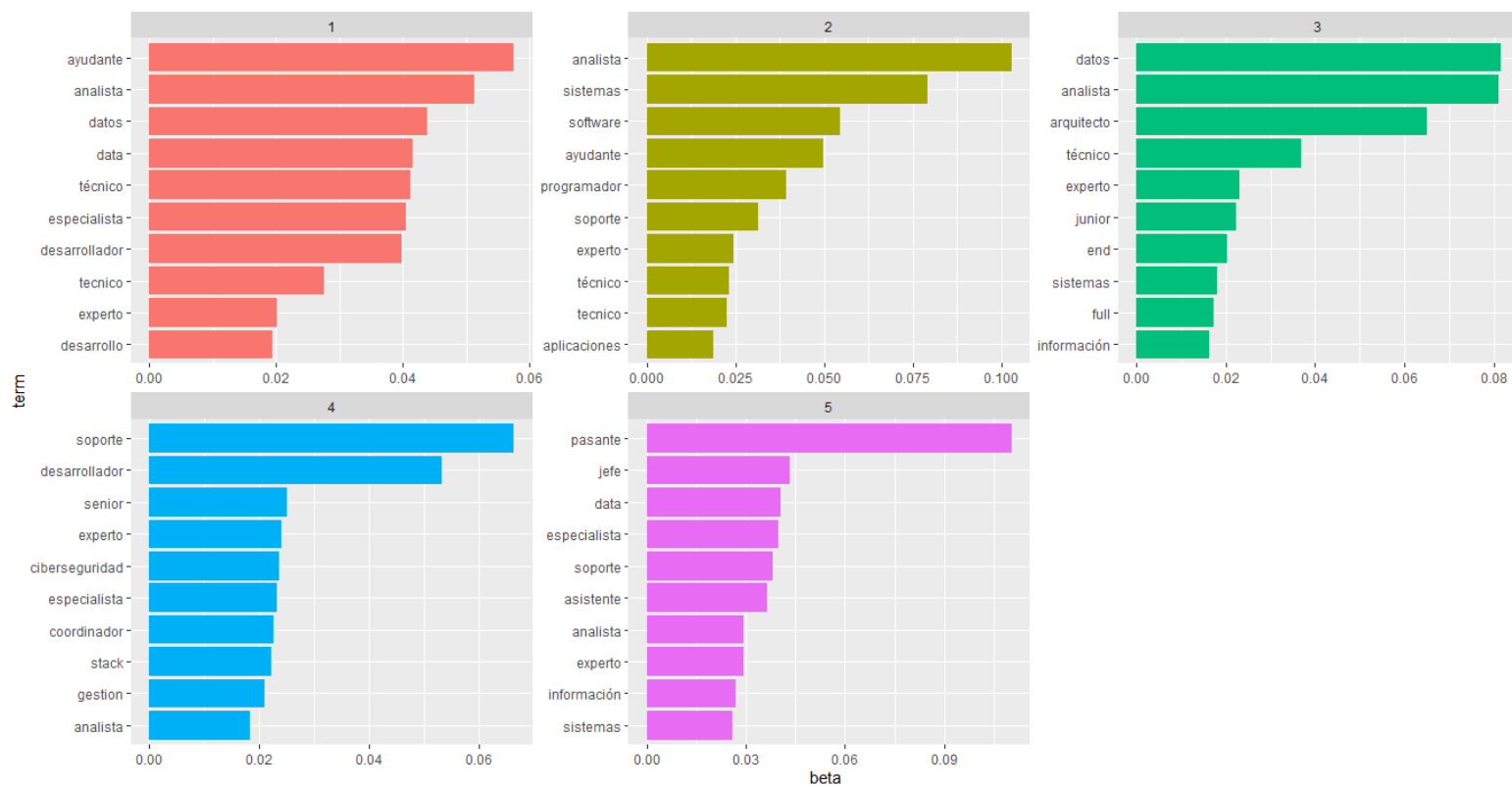
Tópico 4: Desarrollador, testeo de riesgo,

Tópico 5: Gobierno de datos, tecnología de la información y soporte técnico

Tópico 6: Analista de datos, instalaciones de fibra óptica

Tópico 7: Programación, electricidad, conductor, fibra óptica

Tópico 8: Auditoria y fiscalización informática, arquitectura de base de datos, ciberseguridad



Tópico 1: Análisis de datos, desarrollador,

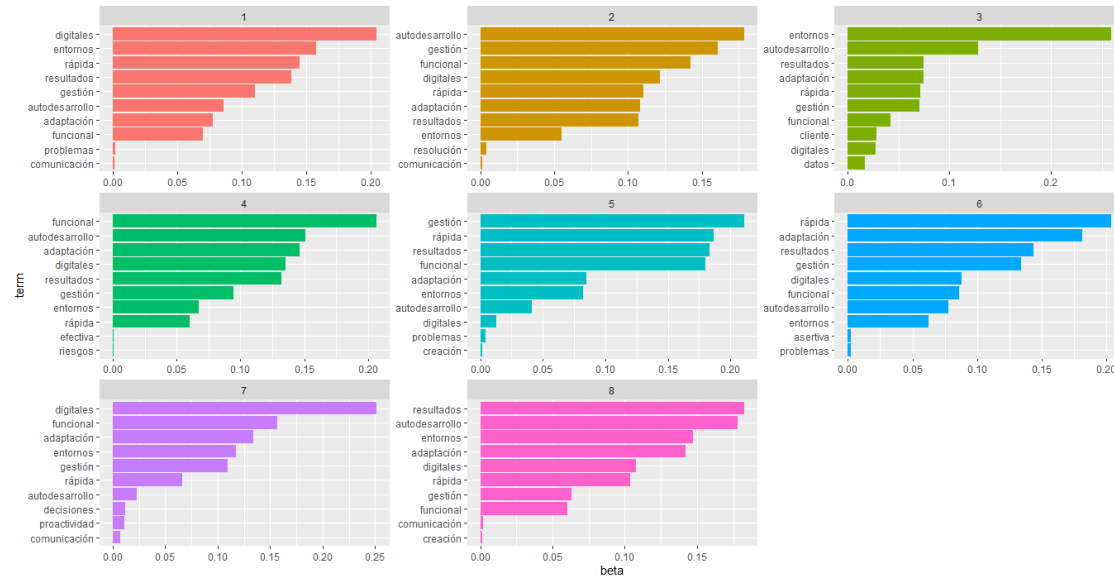
Tópico 2: Programación de software, soporte técnico, aplicaciones web

Tópico 3: Arquitectura de datos, análisis de datos

Tópico 4: Ciberseguridad

Tópico 5: Análisis de datos

Anexo 8: Creación de tópicos mediante Localización Latente de Dirichlet para habilidades blandas



Tópico 1: Autodesarrollo, adaptación rápida, comunicación, funcional

Tópico 2: Resultados

Tópico 3: Atención al cliente

Tópico 4: Desarrollo en entornos digitales

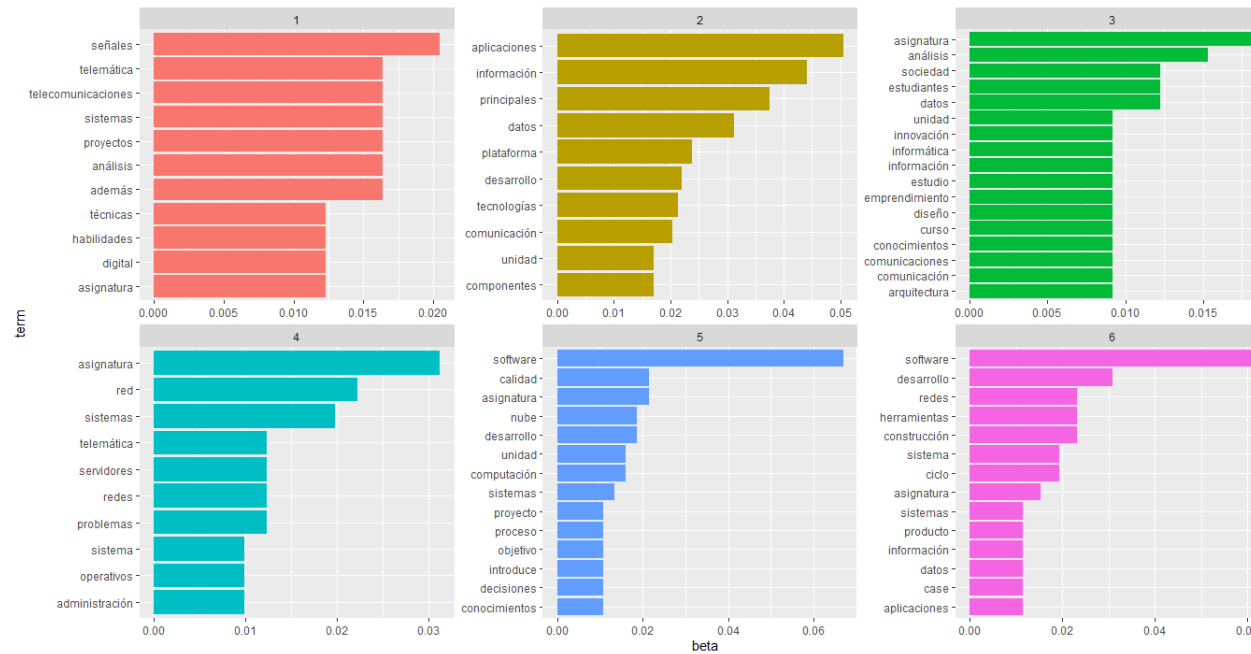
Tópico 5: Resolución de problemas

Tópico 6: Asertivo

Tópico 7: Proactividad, comunicación

Tópico 8: Gestión de las emociones, trabajo bajo presión

Anexo 9: Creación de tópicos mediante LDA a las descripciones de los planes analíticos



Tópico 1: Señales, telecomunicaciones y telemática, habilidades digitales

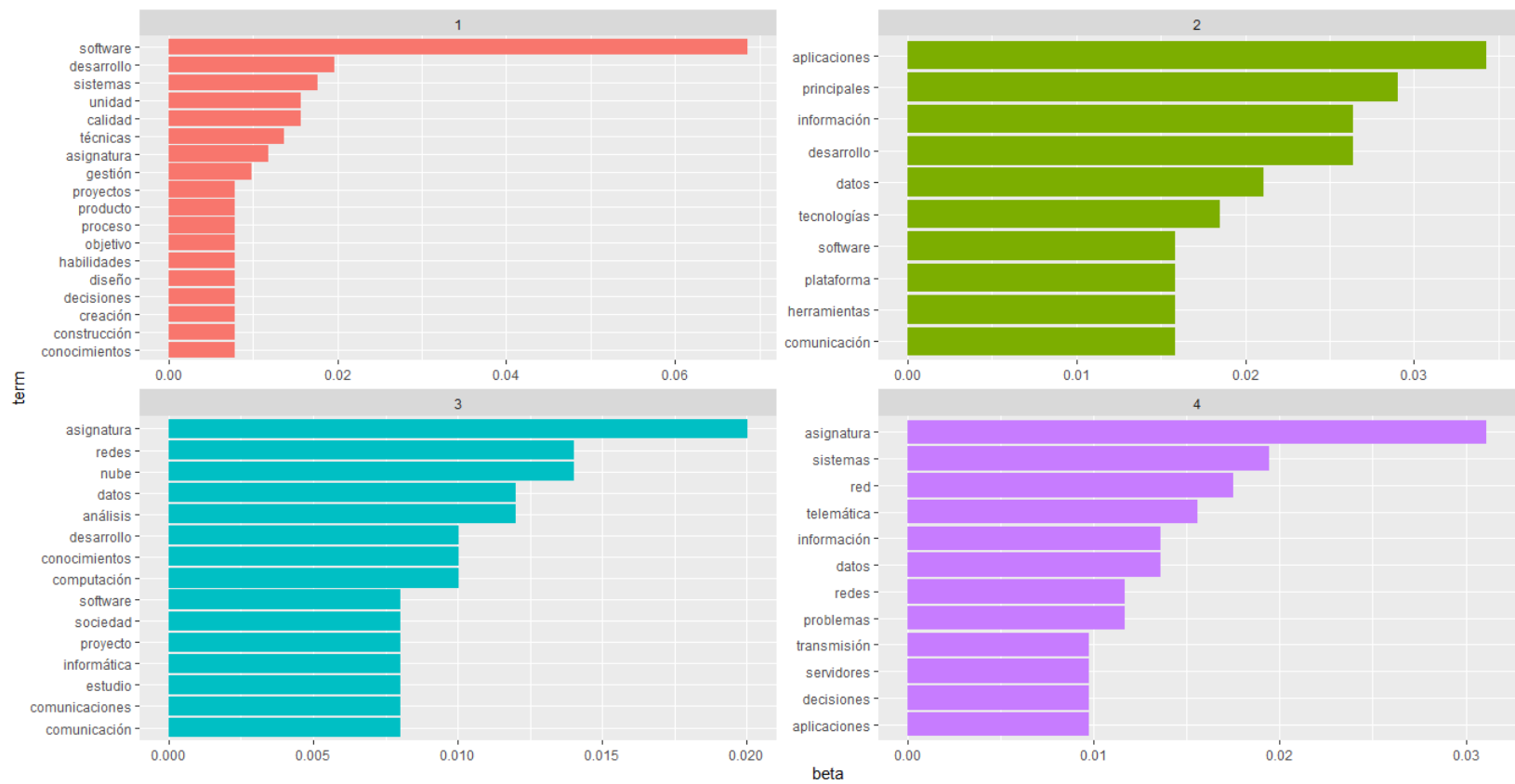
Tópico 2: Desarrollo de tecnologías y aplicaciones y plataformas, datos

Tópico 3: Análisis de la sociedad, innovación, arquitectura de los datos, emprendimientos

Tópico 4: Redes, telemática, servidores, sistemas operativos, administración

Tópico 5: Calidad de software, nube

Tópico 6: Desarrollo de software, construcción y tendido de redes, datos



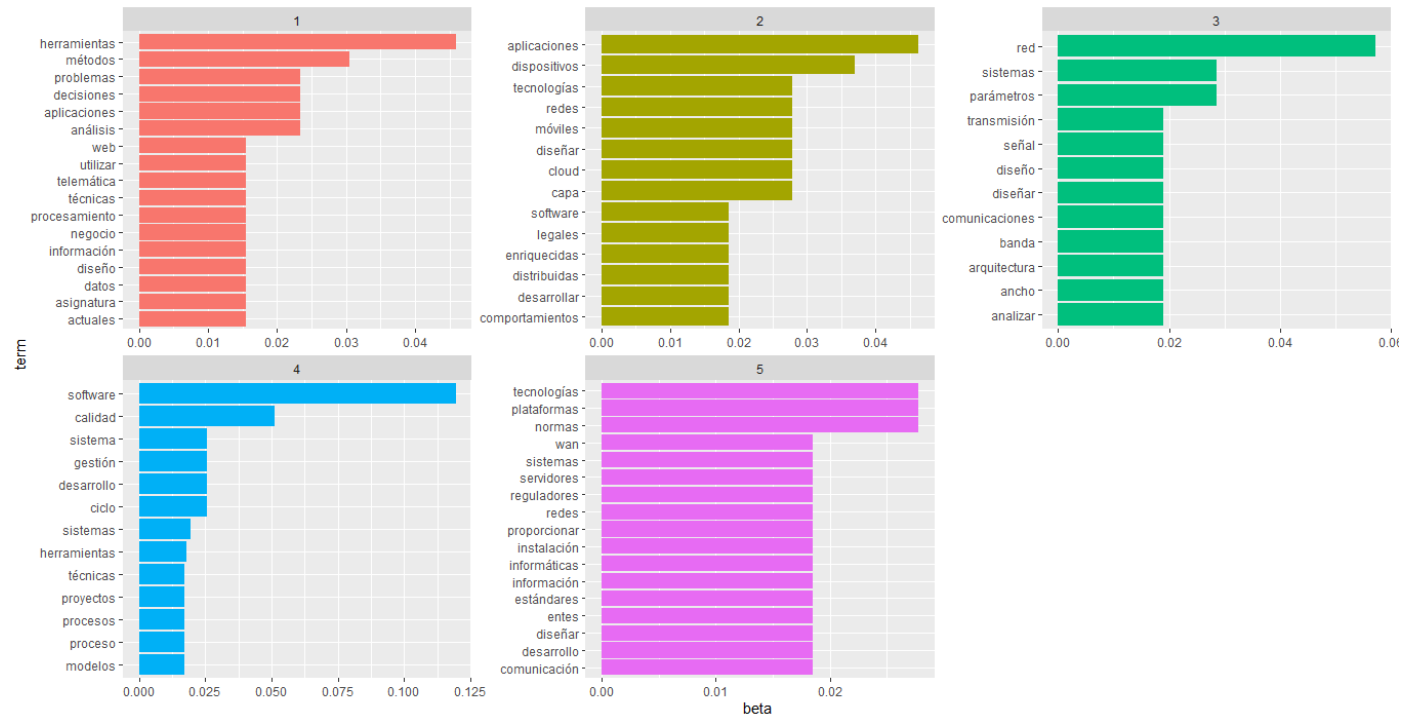
Tópico 1: Calidad y desarrollo de software, gestión de proyectos,

Tópico 2: Plataforma cloud, datos, desarrollo de software

Tópico 3: Redes, nube, análisis de datos, software y comunicación,

Tópico 4: Redes, telemática, transmisión de datos, servidores y aplicaciones

Anexo 10: Creación de tópicos mediante LDA a los objetivos generales de los planes analíticos de ambas carreras



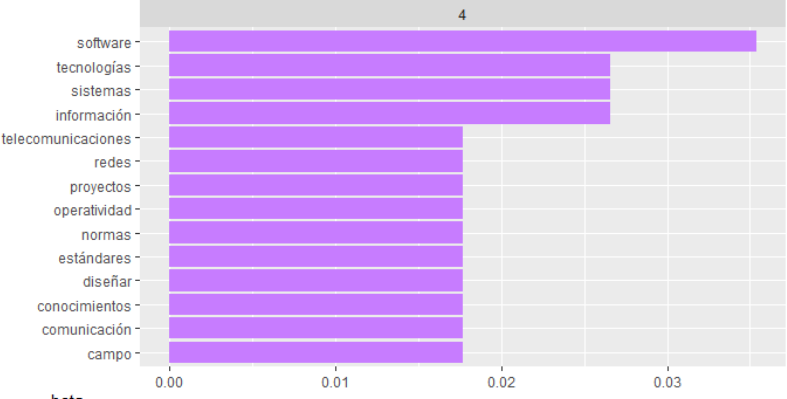
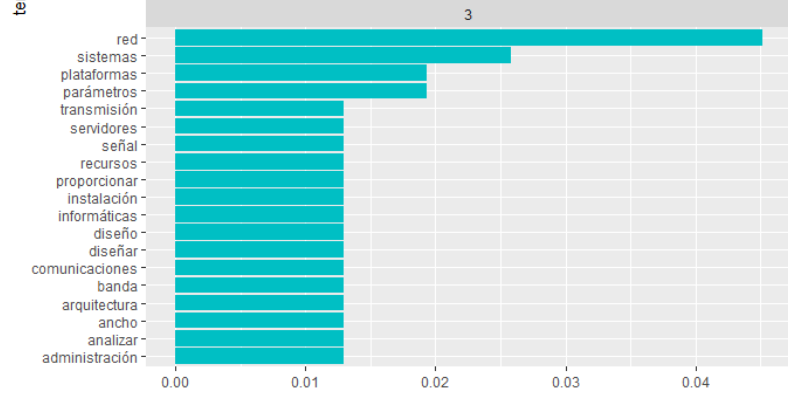
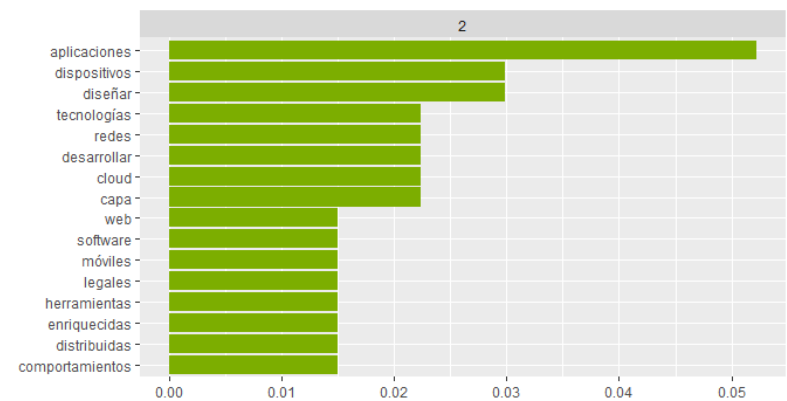
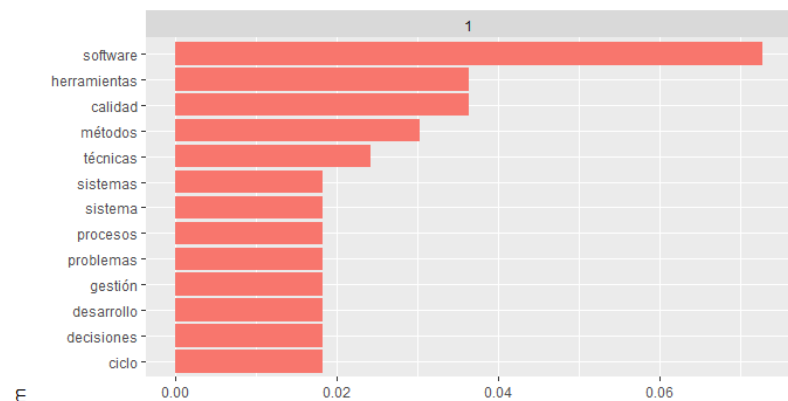
Tópico 1: Análisis y procesamiento de los datos, aplicaciones web, problemas actuales, telemática,

Tópico 2: Aplicaciones móviles, cloud, redes

Tópico 3: Redes, banda ancha, señales, transmisiones, comunicaciones

Tópico 4: Calidad y gestión de software

Tópico 5: Redes WAN, servidores, instalaciones, comunicación



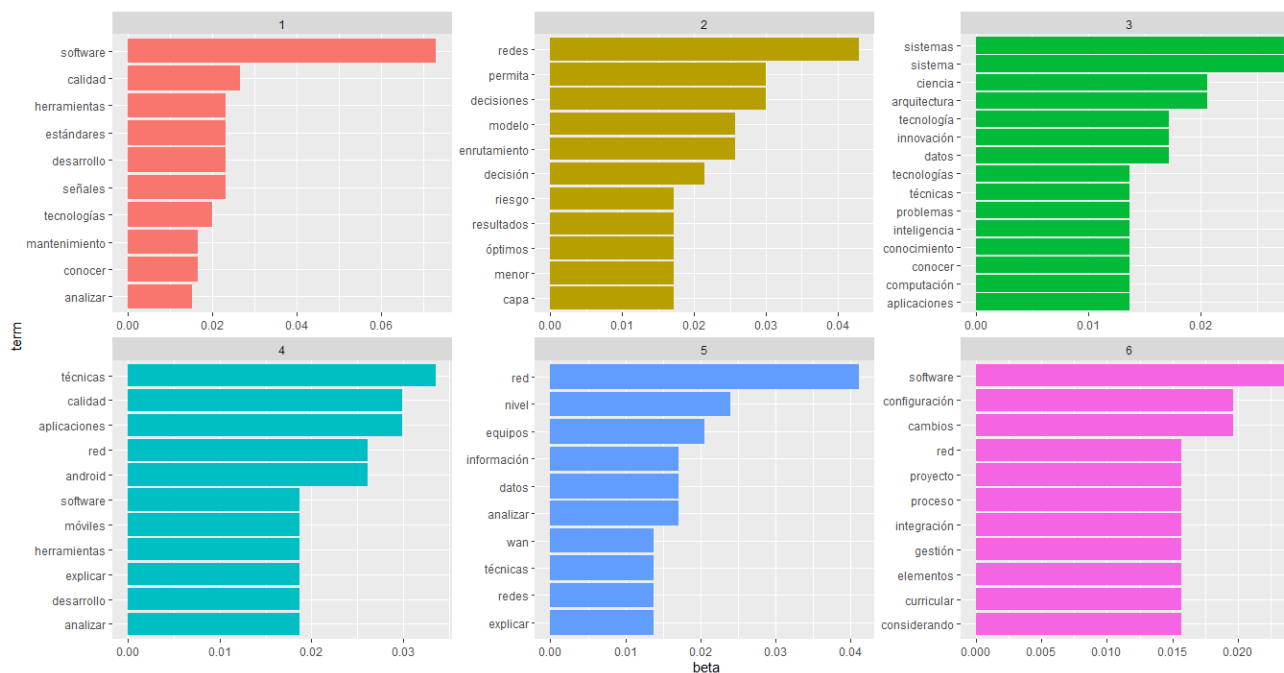
Tópico 1: Desarrollo y gestión de software y calidad

Tópico 2: Aplicaciones y dispositivos, nube, redes, comportamientos legales

Tópico 3: Banda ancha, tendido de redes, redes, señales, transmisiones, instalaciones

Tópico 4: Redes telecomunicaciones, aplicaciones en el campo

Anexo 11: Creación de tópicos mediante LDA para los objetivos específicos de los planes analíticos de cada carrera



Tópico 1: Mantenimiento y calidad de software, señales

Tópico 2: Enrutamiento, modelos óptimos

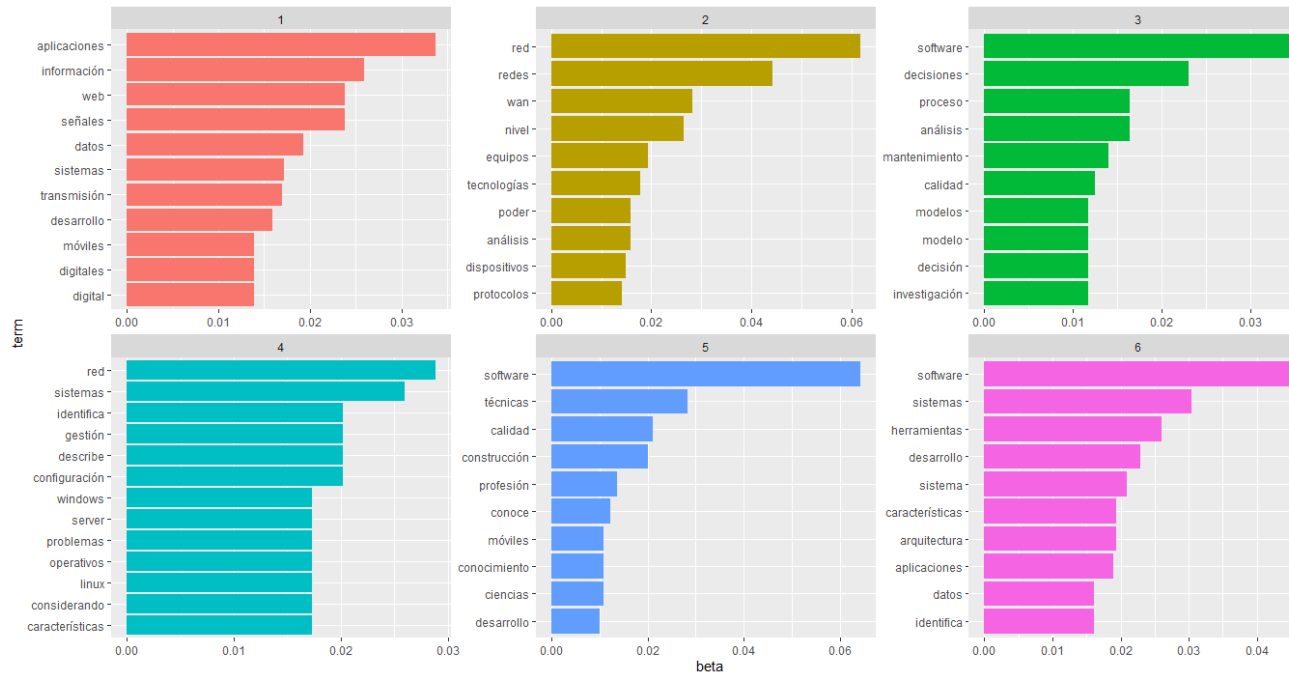
Tópico 3: Arquitectura de base de datos, técnicas innovadoras, aplicaciones

Tópico 4: Aplicaciones, sistemas Android, móviles, análisis de la calidad de software

Tópico 5: Técnico de análisis de datos, redes WAN

Tópico 6: Configuración de software, redes

Anexo 12: Creación de tópicos mediante LDA a los resultados de los planes analíticos de ambas carreras



Tópico 1: Aplicaciones web, transmisión de señales, aplicaciones móviles y digitales

Tópico 2: Redes WAN, protocolos, dispositivos

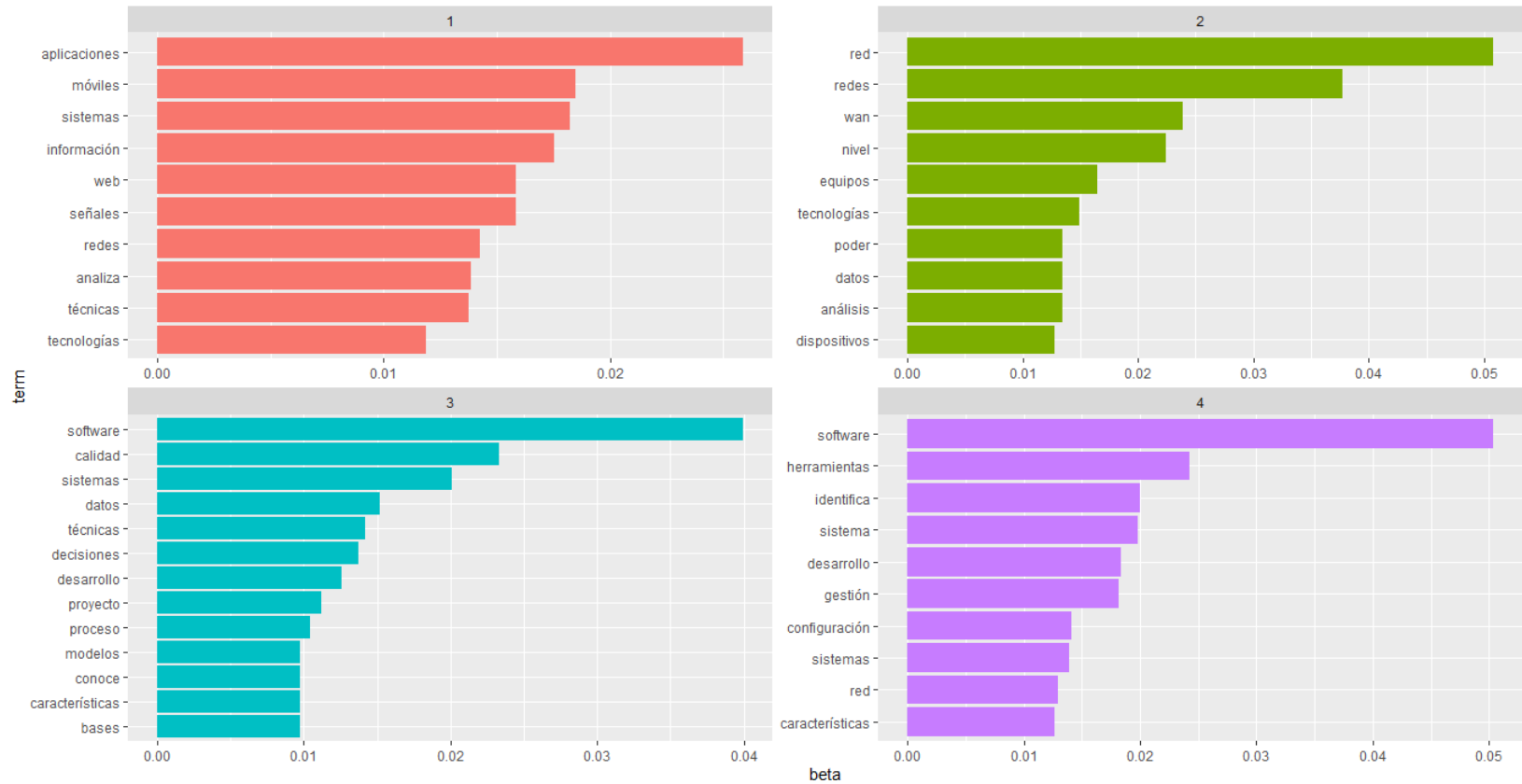
Tópico 3: Modelos de machine learning para la toma de decisiones, análisis de datos

Tópico 4: Linux, servidores, configuración, Windows, sistemas operativos,

Tópico 5: Aplicaciones móviles,

Tópico 6: Arquitectura de base de datos, herramientas, sistemas y software

Anexo 13: Creación de tópicos mediante LDA a los temas y subtemas de los planes analíticos de ambas carreras




Tópico 1: Aplicaciones móviles, tendido de redes,

Tópico 2: Redes WAN, dispositivos, Analítica de datos

Tópico 3: Calidad de software, modelos de machine learning, bases de datos

Tópico 4: Redes, gestión de sistemas

Anexo 14: Ejemplo de un anuncio de empleo del portal Hiring Room



ARQUITECTO SOLUCIONES EXPERTO

📍 Quito, Pichincha, Ecuador
🏢 Tecnología, Sistemas y Telecomunicaciones/Tecnología / Sistemas

🕒 Full-time 🏠 Presencial

[Postularse](#)

NUEVO Hace 21 horas

Descripción del puesto

Diseñar y evaluar las soluciones tecnológicas actuales y propuestas a fin de garantizar una arquitectura y entorno tecnológico simple, adaptable y escalable a las necesidades de la institución, de acuerdo con políticas y procedimientos institucionales, y lineamientos de su línea de supervisión.

Requisitos

EDUCACIÓN

- Cuarto nivel en Tecnología Informática o Telecomunicaciones (Físicas o Matemáticas también serán evaluados si disponen de experiencia contrastable en TI)
- Certificaciones en algún ámbito de Tecnología de Información (Físicas o Matemáticas también serán evaluados si disponen de experiencia contrastable en TI)

EXPERIENCIA

- Experiencia previa en roles de liderazgo o gestión de equipos técnicos y mentoría y desarrollo de otros arquitectos y miembros del equipo técnico.
- Experiencia en Arquitectura de Software, comprensión de patrones de diseño, principios de diseño y capacidad de diseñar sistemas escalables y eficientes. Deseable haber trabajado como desarrollador de software en el pasado o tener una comprensión profunda del desarrollo de software.
- Amplio Conocimiento Tecnológico, manejo de tendencias y tecnologías en el mundo del desarrollo de software y la arquitectura de TI.
- Gerencia de procesos de planificación, control presupuestario, evaluación y dirección de proyectos de Tecnologías de Información.
- Gestión de proyectos complejos de Tecnología de Información Gestión de procesos de cambios y mejoras de Tecnología de Información.
- Manejo y Administración de Proveedores

CONOCIMIENTOS

- Componentes técnicos de los dominios/productos/jornadas desarrolladas.
- Líder técnico y desarrollador experimentado con visión clara acerca de las últimas tecnologías y tendencias.
- Conceptos de arquitectura empresarial.
- Deseable conocimiento en marco de trabajo togaf versión 9.1.
- Modelamiento de Arquitectura de software en formato C4
- Patrones Arquitectónicos: MicroServicios, BFF, EDA, SAGA, Micro-Kernel, CQRS.
- Arquitectura de front-end: MicrofrontEnd, Self-contained Systems
- Metodología de Trabajo: Agile
- Deseable certificaciones en Azure Solutions Architect o AWS Solutions Architect.
- Conocimiento práctico de Infraestructura en la nube (AWS, Azure, Google Cloud) y soluciones SaaS.
- Sólidos conocimientos de lenguajes de programación modernos y frameworks.
- Comprensión de la seguridad cibernética y mejores prácticas de protección de datos.
- Dominio de bases de datos, tanto SQL como NoSQL.
- Habilidad para trabajar con herramientas de CI/CD y control de versiones. (Azure DevOps)

Nosotros

En Banco Pichincha nos estamos transformando; por eso, queremos incorporar el mejor talento a nuestro equipo. Buscamos profesionales amantes de la innovación, la tecnología y el mundo digital que estén dispuestos a asumir nuevos desafíos. Nuestros 117 años de trayectoria nos ratifican que contar con personas comprometidas y con un alto desempeño es imprescindible para lograr nuestros objetivos. ¡Tú puedes ser el próximo! El esfuerzo, la responsabilidad, la coherencia y la trascendencia son valores que marcan la labor diaria de los 5.400 colaboradores que formamos parte de la mayor institución financiera del país. Empezamos a escribir la primera página de nuestra historia en 1906. Hoy contamos con filiales en seis países. En cada uno de esos mercados, defendemos la idea de una banca con propósito porque la tecnología, por sí misma, no es suficiente para favorecer la inclusión financiera. Nuestro liderazgo se afianza más en el nuevo humanismo y la ética empresarial, dos principios que contribuyen a mejorar la vida de nuestros clientes y a aportar al desarrollo de la sociedad.

[Portal de empleos](#) [Compartir](#)

Otras vacantes

DISEÑADOR UX / WRITING

Quito, Pichincha, Ecuador
Diseño / Diseño Web

OFICIAL COMERCIAL Y SERVICIOS - AG FICOA

Ambato, Tungurahua, Ecuador
Comercial, Ventas y Negocios / Comercial

ADMINISTRADOR VOLANTE - GUAYAQUIL

Guayaquil, Guayas, Ecuador
Comercial, Ventas y Negocios / Comercial

92

Anexo 15: Ejemplo de plan analítico de la carrera de Software



**Universidad
Técnica Estatal de Quevedo**
La primera Universidad Agropecuaria del País



PROGRAMA ANALÍTICO

1. DATOS GENERALES

1.1 CARRERA: SOFTWARE (REDISEÑO)	1.2 NIVEL O CICLO: 10MO NIVEL	1.3 MODALIDAD:
1.4 ASIGNATURA: PROGRAMACIÓN EN LA NUBE		1.5 CÓDIGO: ISR-1004
1.6 CRÉDITOS: 3.0	1.7 HORAS PRESENCIALES TEÓRICAS Y PRÁCTICAS: 96.0	1.8 HORAS DE APRENDIZAJE AUTÓNOMO: 48.0
1.9 PRERREQUISITOS: HERRAMIENTAS DE PROGRAMACIÓN - IVO NIVEL		1.10 CORREQUISITOS:
1.11 EJE DE FORMACIÓN: FORMACION PROFESIONAL		

2. DESCRIPCIÓN DE LA ASIGNATURA

La asignatura de Programación en la Nube introduce en los conceptos básicos de la computación en la nube. El estudiante podrá adquirir conocimientos fundamentales necesarios para comprender la computación en la nube desde una perspectiva empresarial y tecnológica. Se abordarán la definición y características esenciales de la computación en la nube, su historia, el caso de negocios de la computación en la nube y los usos de tecnologías basadas en Inteligencia Artificial que permite la nube, así como también los desafíos y las consideraciones éticas de su implementación. Se presentarán algunos de los proveedores de servicios más destacados de la actualidad (por ejemplo, AWS, Google, IBM, Microsoft, entre otros) los servicios que ofrecen y se examinarán algunos estudios de casos de computación en la nube en los distintos sectores verticales de la industria. Tiene como propósito desarrollar en el estudiante la capacidad de describir las arquitecturas y herramientas para desplegar plataformas de Cloud Computing propias o basadas en proveedores.

3. OBJETIVOS DE LA ASIGNATURA

3.1. Objetivo general
Gestionar aplicaciones cloud que permitan satisfacer las necesidades y requerimientos de la organización.

3.2. Objetivos específicos

1. Identificar las características de los diferentes modelos de servicio cloud existentes
2. Clasificar los diferentes componentes de una arquitectura Cloud Computing y los servicios que presta
3. Aplicar principios y herramientas para el desarrollo de aplicaciones en las nubes usando tecnologías emergentes como la Inteligencia Artificial
4. Conocer los principios y fundamentos de la Computación en la Nube

4. RESULTADOS DE APRENDIZAJE DE LA ASIGNATURA

- Identifica oportunidades para elaborar aplicaciones en la nube.
- Establece propuestas metodológicas para el Análisis, Diseño y Desarrollo de proyectos del Software como Servicio.
- Aplica principios y herramientas para el desarrollo de aplicaciones en las nubes y el despliegue de los servicios

5. RESULTADOS DE APRENDIZAJE DE LA CARRERA

1.- Diseña un software mediante la definición de la arquitectura, componentes, interfaces y otras características de un sistema informático que permita su adaptación y crecimiento en función del desarrollo-organización y para la gestión eficiente

6. CONTENIDOS PROGRAMÁTICOS CON RESULTADOS DE APRENDIZAJE

6.1. RESULTADO DE APRENDIZAJE 1: Conoce los principios y fundamentos de la Computación en la Nube 46573

UNIDAD 1: FUNDAMENTOS DE LA COMPUTACIÓN EN LA NUBE

TEMA 1: ENCUADRE ACADÉMICO

TEMA 2: Introducción al Cloud Computing

TEMA 3: Historia y Evolución del Cloud Computing

www.uteq.edu.ec



Campus "Ingeniero Manuel Agustín Haz Álvarez", Av. Quito km. 1 1/2 vía a Santo Domingo de los Tsáchilas
Tel: (593) 5370-2220 - info@uteq.edu.ec - www.uteq.edu.ec
QUEVEDO - LOS RÍOS - ECUADOR



TEMA 4: Principales proveedores de Cloud Computing y los servicios que ofrecen
TEMA 5: Inteligencia Artificial en la Nube

6.2. RESULTADO DE APRENDIZAJE 2: Identifica las características de los diferentes modelos de servicio cloud existentes 46575

UNIDAD 2: MODELOS DE CLOUD COMPUTING

TEMA 1: Visión general de los Modelos de Servicios en la Nube
TEMA 2: IaaS - Infraestructura como Servicio
TEMA 3: PaaS - Plataforma como Servicio
TEMA 4: SaaS - Software como Servicio
TEMA 5: AaaS - Inteligencia Artificial como Servicio

6.3. RESULTADO DE APRENDIZAJE 3: Clasifica los diferentes componentes de una arquitectura Cloud Computing y los servicios que presta 46574

UNIDAD 3: COMPONENTE DE LA ARQUITECTURA CLOUD COMPUTING

TEMA 1: Visión General de la Infraestructura de la Nube
TEMA 2: Virtualización y Máquinas Virtuales
SUBTEMA 1: Tipos de Máquinas Virtuales
TEMA 3: Servidores Bare Metal
TEMA 4: Contenedores

6.4. RESULTADO DE APRENDIZAJE 4: Aplica principios y herramientas para el desarrollo de aplicaciones en las nubes usando tecnologías emergentes como la Inteligencia Artificial 46576

UNIDAD 4: Tecnologías emergentes de la Computación en la Nube

TEMA 1: Inteligencia Artificial en la Nube
SUBTEMA 1: El poder de la Inteligencia Artificial en la Nube
SUBTEMA 2: Desafíos y Consideraciones al Implementar tecnologías de IA en la Nube
SUBTEMA 3: Uso ético de la IA en la nube
TEMA 2: Multi-Nube Híbrida
TEMA 3: Microservicios
TEMA 4: Aplicaciones Nativas de la Nube

7. METODOLOGÍA

- Las actividades prácticas se ejecutarán mediante una herramienta de desarrollo de aplicaciones web en el laboratorio de cómputo
- El aprendizaje basado en proyectos se evidenciará con el despliegue de un Software Como Servicio (SaaS) y la implementación de caso reales mediante el uso de otras APIs y Servicios Web
- Los trabajos autónomos implican la realización de trabajos de consulta, exposiciones, tareas individuales y grupales
- La principal herramienta Tecnológica-pedagógica será el Sistema de Gestión Académica (SGA) como la plataforma educativa y administrativa central, donde se presentarán los planes de estudio, asistencias de estudiantes, control de calificaciones y distribución de recursos de aprendizajes y demás elementos que constan en el "Aula Virtual". Además del SGA, se publicarán los recursos en el google drive que se creará para el efecto

8. PROCEDIMIENTO DE EVALUACIÓN

REFERENTES	%	CALIFICACIÓN	DESCRIPCIÓN
Primer Corte	35	Sobre máximo de 10 puntos	Son productos del proceso de evaluaciones frecuentes (los tipos más utilizados son: observación del trabajo de los estudiantes, preguntas orales y escritas, discusiones grupales, entre otros) y parciales (la prueba parcial, el trabajo extracurricular, el encuentro comprobatorio, entre otros).
Segundo Corte	35	Sobre máximo de 10 puntos	Son productos del proceso de evaluaciones frecuentes (los tipos más utilizados son: observación del trabajo de los estudiantes, preguntas orales y escritas, discusiones grupales, entre otros) y parciales (la prueba parcial, el trabajo extracurricular, el encuentro comprobatorio, entre otros).
Examen Final	30	Sobre máximo de 10 puntos	Evaluará, fundamentalmente, los conocimientos y las habilidades adquiridos por el estudiante, y su capacidad para integrarlos, sistematizarlos, aplicarlos y generalizarlos