



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

PROCESO DE POSTULACIÓN Y SELECCIÓN ESTUDIANTIL EN EQUIPOS  
DOCENTES EN LA ESCUELA DE INGENIERÍA Y CIENCIAS DE LA FCFM,  
ANÁLISIS CON PERSPECTIVA DE GÉNERO

MEMORIA PARA OPTAR AL TÍTULO DE  
INGENIERA CIVIL INDUSTRIAL

CAROLINA ANDREA FIGUEROA VALDIVIA

PROFESOR GUÍA:  
SERGIO CELIS GUZMÁN

MIEMBROS DE LA COMISIÓN:  
VÍCTOR PÉREZ VERA  
FERNANDA MELIS JACOB

SANTIAGO DE CHILE  
2023

# Resumen

En la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile (FCFM) existen dos cargos principales de apoyo a la Docencia en la FCFM: Profesores Auxiliares y Ayudantes, quienes desempeñan las actividades que sean designadas por el profesor de cátedra o coordinador del curso, entre las que se encuentran la realización de clases de apoyo, coordinación, revisión y cuidado del equipo, y quienes además deben dar cumplimiento a una serie de requisitos y normas principalmente relacionadas a su desempeño académico y ético. El proceso de llamado, participación en concurso y selección a equipos docentes puede ser realizado como cada unidad docente determine, sin existir un proceso estandarizado a nivel FCFM, sin embargo, el registro formal de estos debe oficializarse mediante las plataformas Ucampus y Ucursos. Pese a que se cuenta con un extenso registro de datos relacionados a la postulación, selección y participación de estudiantes en equipos docentes, estos actualmente no son analizados de forma periódica ni profunda por ningún área de la FCFM.

Dado lo anterior es que el objetivo general del presente trabajo de memoria es analizar los factores que explican la selección de estudiantes en equipos docentes de la Escuela de Ingeniería y Ciencias de la FCFM de la Universidad de Chile a través de modelos clasificadores. De tal forma, se busca que los resultados de la investigación problematicen la brecha de género que pudiese existir en estos procesos académicos formativos. Para ello se analizarán las postulaciones efectivas y las selecciones de estudiantes a los equipos docentes, para entender sesgos, barreras y oportunidades en los procesos de postulación y selección en los equipos docentes.

Para esto se utilizó un conjunto de datos de 5 años que incluyó información sobre la carrera, el género, la edad, la experiencia como equipo docente previa, las notas y los cursos anteriores de los postulantes, así como información sobre si fueron seleccionados o no para el cargo. Se realizó un análisis exploratorio y descriptivo de los datos, así como pruebas de hipótesis y modelos de aprendizaje automático para identificar posibles patrones y relaciones entre las variables. Los resultados indicaron que el género no parecía ser un factor significativo en la selección de los estudiantes, pero se observaron otras características relevantes, como la experiencia previa como equipo docente y la carrera de estudio.

En general, los resultados sugieren que se necesitan más investigaciones y medidas para fomentar la igualdad de género en la selección de estudiantes para equipos docentes en carreras de ingeniería. Además, se destaca la importancia de considerar otras variables más allá del género en la selección de estudiantes, como la experiencia previa y el rendimiento académico, para garantizar una selección justa y equitativa.

*Dedicada a mi familia, que siempre cree en mí.*

# Agradecimientos

De tan solo pensar en el fin de esta etapa no puedo evitar sentir mucha emoción, recuerdo estar en segundo medio y empezar a plantearme a mí misma que quería ser ingeniera. Desde ese día fue un camino muy duro, en el que me tuve que enfrentar a muchas cosas, pero habían personas en mi vida que nunca dudaron de mí, y que sin duda gracias a ellos nunca me rendí. En primer lugar mis papás Edith y Cristian, quienes desde pequeña me inculcaron tanto la importancia de mi educación como de mi felicidad. Gracias papitos, por cada palabra de aliento, porque no me dejaron dudar de mí misma y por sobre todo siempre confiar en mis decisiones, nunca escuche palabras de dudas de ustedes, eso y el gran amor que me entregaron me ayudo a ser quien soy. A mi hermanito, quien cada día me enorgullece más. No puedo no agradecer además a mis tíos Hochi y Nancy, quienes siempre me apoyaron incluso hasta el día de hoy, con sus palabras sabias y el gran amor que me han tenido siempre, mis papis postizos siempre los tengo en mi corazón. Además a mi abuelita, quien pese a la distancia siempre se hizo presente en mi vida universitaria.

Desde que decidí emprender este camino también lo vivieron conmigo mis amigas del colegio, quienes siempre se reían por mi obsesión con la Chile, pero dentro de esas risas existió también el apoyo incondicional Javi, Cote, Coni, Cata, Domi, Dani, Iva, Consu, Sofi gracias por ser un pilar tan fundamental en mi vida tantos años, las amo. Dentro de este camino también tuve palabras de alientos y muy buenos momentos con mis amigos que siempre me sacan sonrisas Willy, Cele, Diego, Max, Jorge, Oso, Jabi, Cril, Pipe, Pato. Gracias mi Gang, los tkm infinito.

No puedo no agradecerle a las personas que desde que entre a la u fueron mis mejores amigos y compañía, Gonzalo y por sobre todo Tomi, fuiste un gran apoyo y una maravillosa persona en mi época universitaria y por eso siempre estaré agradecida. Quiero agradecer además a mis niños de Quilpue, mis amigos de taller, mis profesores y a quienes sin duda marcaron mi último periodo universitario, mis niños Angie, Jose, Nati, Agus, Marquitos, Diego, Eybie, Pipe, Barbi, Nico, Benja, Marquitos, Clau, Pablito, Tere y Pablito Woo. Y para terminar pero no por eso menos importante, tengo que agradecerle a las personas que el día de hoy se transformaron en mi familia elegida, Yerko, Chiki, Ani y Nicole, gracias por enseñarme tanto lo que es el amor incondicional y por apoyarme en cada momento. Nicole no podía no hacer un apartado solo para ti, gracias por llegar a mi vida, por estar en los buenos y malos momentos, yo sé que viviremos muchas aventuras más. Te amo mucho.

Y qué más puedo decir, gracias a la vida por todas las oportunidades que me ha dado.

# Tabla de Contenido

<b>1. Introducción</b>	<b>1</b>
1.1. Antecedentes Generales . . . . .	1
1.2. Pregunta de Investigación . . . . .	4
1.3. Objetivos . . . . .	4
1.4. Alcances . . . . .	5
<b>2. Marco de Referencia</b>	<b>6</b>
2.1. Perspectiva de Género . . . . .	6
2.2. Brechas de Género y su importancia transversal . . . . .	7
2.3. STEM y Sesgos . . . . .	7
2.4. Selección de cargos docentes en la Facultad de Ciencias Físicas y Matemáticas	10
<b>3. Diseño Metodológico</b>	<b>13</b>
3.1. Descubrimiento de conocimiento en base de datos - Metodología KDD . . . . .	13
3.2. Análisis exploratorio de datos (EDA) . . . . .	15
3.3. Aprendizaje de máquina supervisados . . . . .	15
3.3.1. Métodos de conjunto . . . . .	16
3.3.2. Random Forest . . . . .	16
3.3.3. Support Vector Machine . . . . .	16
3.3.4. XGBoost . . . . .	16
3.3.5. Regresión logística . . . . .	16
3.3.6. Herramientas de evaluación de los modelos . . . . .	17

3.4. Datos . . . . .	17
<b>4. Desarrollo metodológico</b>	<b>19</b>
4.1. Análisis Exploratorio y Descriptivo de Datos . . . . .	19
4.1.1. Equipos docentes vs proporción de género según carrera a través de los años . . . . .	19
4.1.2. Postulantes vs proporción de género según carrera a través de los años	21
4.1.3. Proporción de postulantes vs Aceptados . . . . .	23
4.1.4. Rendimiento académico del postulante según género . . . . .	25
4.1.5. Experiencia académica según género . . . . .	27
4.1.6. Experiencia académica según género . . . . .	28
4.1.7. Gráficos de distribución . . . . .	29
4.1.8. Matriz de Correlación . . . . .	33
4.1.9. ANOVA . . . . .	34
4.1.10. Test de hipótesis . . . . .	34
<b>5. Desarrollo de los modelos</b>	<b>36</b>
5.1. Definición de variables . . . . .	36
5.2. Estructura y optimización de los modelos . . . . .	37
5.3. Entrenamiento y resultados . . . . .	38
<b>6. Discusión de resultados</b>	<b>42</b>
<b>7. Conclusiones</b>	<b>45</b>
7.1. Conclusiones generales del trabajo . . . . .	45
7.2. Recomendaciones para la facultad y trabajos futuros . . . . .	46
<b>Bibliografía</b>	<b>50</b>

# Índice de Tablas

3.1. Lista de variables y su descripción . . . . .	18
4.1. Resultados de la ANOVA para las variables de interés. Fuente: elaboración propia . . . . .	34
4.2. Tabla de contingencia y resultados del test de hipótesis. Fuente: Elaboración propia . . . . .	35
5.1. Descripción de las variables. Fuente: Elaboración Propia . . . . .	37
5.2. Matriz de confusión SVM. Fuente: Elaboración propia. . . . .	38
5.4. Matriz de confusión Random Forest. Fuente: Elaboración propia. . . . .	38
5.3. Matriz de confusión XGB. Fuente: Elaboración propia. . . . .	39
5.5. Matriz de Regresión Logística. Fuente: Elaboración propia. . . . .	39
5.6. Importancia de las variables en el modelo SVM. Fuente: Elaboración propia. . . . .	39
5.7. Importancia de las variables en el modelo XGBoost. Fuente: Elaboración propia. . . . .	40
5.8. Importancia de las variables en el modelo Random Forest. Fuente: Elaboración propia. . . . .	40
5.9. Coeficientes de las variables en el modelo de regresión logística. Fuente: Elaboración propia. . . . .	41

# Índice de Ilustraciones

1.1. Evolución del porcentaje de mujeres auxiliares en el departamento de ingeniería civil industrial . . . . .	3
3.1. El proceso KDD . . . . .	14
4.1. Evolución de mujeres seleccionadas y proporción de mujeres según carrera . .	20
4.2. Evolución de hombres seleccionados y proporción de hombres según carrera .	21
4.3. Evolución de mujeres postulantes y proporción de mujeres según carrera . . .	22
4.4. Evolución de hombres postulantes y proporción de hombres según carrera . .	23
4.5. Evolución de mujeres postulantes vs aceptadas según carrera . . . . .	24
4.6. Evolución de hombres postulantes vs aceptados según carrera . . . . .	25
4.7. Rendimiento académico de estudiantes aceptados y rechazados por género para cada carrera . . . . .	26
4.8. Experiencia previa promedio en cargos docentes de estudiantes aceptados y rechazados por género para cada carrera . . . . .	27
4.9. Cantidad de postulantes que tuvieron un profesor auxiliar de su mismo género al cursar el ramo, por carrera . . . . .	28
4.10. Distribución de la edad al momento de postular para aceptados . . . . .	29
4.11. Distribución de edad al postular por género . . . . .	29
4.12. Distribución de la nota de aprobación para estudiantes aceptados por género	30
4.13. Distribución de experiencia como ED para estudiantes aceptados . . . . .	30
4.14. Distribución de edad al postular para estudiantes aceptados y rechazados . .	31
4.15. : Distribución de rendimiento académico para estudiantes aceptados y rechazados	31
4.16. Distribución de experiencia como ED para estudiantes aceptados y rechazados	32



4.17. Distribución de proporción de estudiantes del mismo género para estudiantes aceptados y rechazados . . . . .	32
4.18. Mapa de calor de la matriz de correlación entre variables numéricas . . . . .	33

# Capítulo 1

## Introducción

### 1.1. Antecedentes Generales

La brecha de género es una realidad que lamentablemente no se ha visto superada a pesar de los diversos esfuerzos tanto a nivel institucional estatal como a nivel institucional privado. Esta arista se ha hecho presente dentro de los 17 Objetivos de Desarrollo Sostenible (ODS), específicamente en los puntos 4 y 5, los que tratan temas de educación e igualdad de género, respectivamente (Jorquera 2020). Es aquí donde el rol de la ciencia y tecnología es clave; la UNESCO sostiene que por medio de las disciplinas STEM , (Ciencias Básicas, Tecnología, Ingeniería y Matemáticas. “. . . ) el cumplimiento de los objetivos de la Agenda 2030 para el Desarrollo sostenible está más cerca, pues son las que pueden entregar a quienes la estudian, los conocimientos, las habilidades, las actitudes y las conductas necesarias para crear sociedades inclusivas y sostenibles (UNESCO 2016 en Jorquera 2020).

Pese a este beneficio social, a nivel mundial en promedio cada 30 por ciento varones que se gradúan de la educación superior pertenecen a un programa formativo STEM, sólo el 16 por ciento de mujeres egresadas corresponden a programas educativos de este tipo. (WEF, 2016). Si se analiza en primer lugar en el caso latinoamericano, más específicamente en el país vecino de Chile, Argentina, según los datos del CIPEEC (Centro de Implementación de Políticas Públicas para la Equidad y el Crecimiento) seis de cada 10 estudiantes universitarios son mujeres, pero si se analiza de forma específica, de ellas solo el 25 % estudia ingeniería y Ciencias Aplicadas, y apenas un 15 % se inscriben en carreras de programación. (Braga, 2021). Si se observa el caso de Chile desde el año 2011 hasta la fecha el porcentaje de ingreso a carreras STEM es cercano al 20 % siendo la cifra más alta registrada de un 23 % en el año 2013 (Cuevas, 2022).

Según el libro “Del biombo a la cátedra : igualdad de oportunidades de género en la Universidad de Chile” (2014), el problema de la segregación horizontal o la desigualdad de hombres y mujeres en las distintas áreas del conocimiento en las instituciones se presenta tanto en la composición del estudiantado como en el cuerpo académico, además se menciona la existencia de una segregación vertical, que indicaría que las mujeres tienen una presencia cada vez menos significativa en la medida en que se avanza en la pirámide de poder y

autoridad. La importancia de esto radica no solo en los involucrados, sino que también en el desempeño académico de estudiantes independiente de su sexo, como también en los procesos que definen la carrera académica, ya que imponen y valoran de manera desigual los desempeños de cada grupo, además afecta en su posterior desempeño profesional en ambos casos, por ejemplo, los hombres egresados de carreras feminizadas tendrían problemas para ser contratados. (Andrade, 2013).

Según la directora del Área Género, Sociedad y Políticas de la Facultad Latinoamericana de Ciencias Sociales (FLACSO-Argentina) y de la Cátedra Regional UNESCO Mujer, Ciencia y Tecnología en América Latina existen factores tales como la falta de modelos de rol, la representación social de la ciencia y la ingeniería y la falta de interacción con ingeniera, los estereotipos sobre el desarrollo profesional, la discriminación explícita o sutil en contextos de estudio y la falta de políticas y prácticas laborales para conciliar trabajo y familia, han impedido que las mujeres se sientan atraídas por estudiar carreras STEM.

En la actualidad en la FCFM se encuentran 2 roles académicos para el estudiante, el profesor/a de cátedra y los equipos docentes compuestos por auxiliares y ayudantes. Estas últimas figuras en general se encargan de asistir en términos de gestión y funcionamiento de cada curso, según le sea asignado por el profesor/a, con actividades tales como docencia auxiliar, coordinación de ayudantes, coordinación de secciones paralelas de un curso, preparación y cuidado de equipo y material para las clases en el caso de los primeros, y por otro lado corrección de pruebas, ejercicios, tareas y laboratorios, asistencias y reclamos en el caso de los segundos. (FCFM, 2018)

Actualmente el proceso de postulación y selección a cargos en equipos docentes para profesor auxiliar o ayudante está a cargo del profesor de cátedra del ramo en cuestión, pero este no considera ningún proceso estandarizado a nivel de facultad. Cuando un/a estudiante desea postular todo el proceso lo debe realizar por Ucampus, en la pestaña de “Postulaciones Docentes”, luego una vez este ha sido seleccionado o rechazado se le notifica a través de la misma plataforma y finalmente quienes son aceptados el profesor lo registra en Ucursos y Ucampus.

Este es un proceso que se realiza en toda la facultad cada semestre, en el que muchos estudiantes seleccionados no sólo apoyan en la educación de sus compañeros, sino que también se relacionan a un nivel mayor con el profesor y el contenido del curso, incluso la Escuela ha declarado la importancia de fortalecer las habilidades de auxiliar y ayudantes para mejorar la docencia en la FCFM, dada la influencia que tiene esta actividad en el desarrollo profesional de los estudiantes. Esto se respalda en la investigación de la memoria de Scarlett Penroz del año 2020 “Análisis de la continuidad de estudios entre pre y post grado de estudiantes de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile, con enfoque de género”, la cual afirma que la participación en los equipos docentes influye positivamente en la realización de algún postgrado en ellos, indistintamente del género del auxiliar. (Penroz, 2020.)

Pese a todo lo mencionado anteriormente, hasta la fecha no se ha analizado con mayor profundidad qué ocurre con este proceso. ¿Quiénes postulan? ¿Con qué criterios son seleccionados? ¿Existe algún sesgo en esta selección?

El presente estudio se plantea investigar la participación de mujeres en puestos representativos dentro de la educación, específicamente en los equipos docentes de la carrera académica de la FCFM de la Universidad de Chile, analizando el comportamiento de ésta a lo largo del tiempo y si presenta alguna brecha en el proceso de selección de este cargo.

Otro punto importante es el aumento de participación de las mujeres en la facultad, considerando que se ha puesto como objetivo en los últimos años aumentar el ingreso de mujeres, logrando entre los años 2000 y 2019 una participación de 22 % de ingreso a pregrado, y que además el estamento docente alcanza un 17 % de participación femenina según datos de la DDG. Es lógico pensar entonces que al aumentar esto, la representatividad femenina en los cargos docentes también sea creciente, y esto es algo que también se analizará más adelante. (Universidad de Chile, sf)

Si bien no se tiene claridad de cómo se comporta la distribución de género en los cargos docentes en la actualidad en la facultad, en la memoria de Florencia Correa (2019) “Análisis de diferencias de género en la evaluación docente de los cursos dictados por el departamento de ingeniería civil industrial” se muestra que entre los años 2010 al 2017 el porcentaje de mujeres que participan como profesoras de cátedra y auxiliares es menor al de los hombres, pese a esto para el año 2017 se logró un porcentaje bastante similar al de la cantidad de mujeres que ingresó al departamento, lo cual daría luces de una representatividad positiva.

Figura 1.1: Evolución del porcentaje de mujeres auxiliares en el departamento de ingeniería civil industrial

<b>Año</b>	<b>Porcentaje de mujeres auxiliares</b>
<b>2010</b>	6%
<b>2011</b>	9%
<b>2012</b>	13%
<b>2013</b>	10%
<b>2014</b>	10%
<b>2015</b>	12%
<b>2016</b>	11%
<b>2017</b>	27%

Fuente: Análisis de diferencias de género en la evaluación docente de los cursos dictados por el Departamento de Ingeniería Civil Industrial, memoria para optar al título de Ingeniera Civil Industrial de Florencia Antonia Correa Campos

Ante todo, lo anteriormente mencionado, nace la oportunidad de estudiar desde la intención de postular a la selección, considerando como información la gran cantidad de datos disponibles en la base de datos de la plataforma U-Campus.

En la primera parte del trabajo se expone la pregunta de investigación y los objetivos, seguido por las principales teorías y líneas de pensamiento respecto a las brechas de género y los sesgos.

Luego, se presenta la metodología que responde a los objetivos planteados en el estudio, en este caso enfocada en el estudio de datos y la utilización de modelos de clasificación.

Posteriormente se discuten los hallazgos de la investigación, lo que produce un análisis

más profundo de quienes postulan y los factores que influyen en la selección de los equipos docentes.

Finalmente se presentan las conclusiones, relacionando los principales resultados de los datos estudiados con el contexto de la universidad y lo señalado por la literatura para finalizar con reflexiones y recomendaciones para futuras investigaciones y la misma facultad.

## 1.2. Pregunta de Investigación

¿Cuáles son los principales factores que influyen en la selección de estudiantes en los equipos docentes de la Escuela de ingeniería y Ciencias de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile?

## 1.3. Objetivos

Se plantea como **objetivo general** analizar los factores que explican la selección de estudiantes en equipos docentes de la Escuela de Ingeniería y Ciencias de la FCFM de la Universidad de Chile a través de modelos clasificadores. De tal forma, se busca que los resultados de la investigación problematicen la brecha de género que pudiese existir en estos procesos académicos formativos.

El supuesto/hipótesis que se buscará comprobar en este trabajo es que gracias a la ampliación de medidas sistemáticas en el área académica y curricular de Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile que han provocado un aumento de estudiantes mujeres en las carreras, se reducirían los sesgos y barreras para que mujeres postulen a cargos docentes, fortaleciendo la equidad de género en el sistema STEM.

Es así que los **objetivos específicos son:**

1. Seleccionar las variables de mayor incidencia sobre la selección de los estudiantes a cargos docentes.
2. Implementar y evaluar distintos modelos de clasificación sobre esta decisión.
3. Determinar posibles sesgos, barreras y oportunidades en el proceso de postulación y selección de los equipos docentes.
4. Identificar las variables que afectan en la decisión de selección de un estudiante para un cargo docente en la FCFM de la Universidad de Chile.

## 1.4. Alcances

Este trabajo espera lograr la comprensión del estado actual del proceso de postulación y selección de los estudiantes a equipos docentes, es importante resaltar que este estudio tendrá distintos alcances debido al tamaño y complejidad de la muestra real del universo estudiado, el cual corresponde a todos los estudiantes de la FCFM, partiendo por los años a estudiar, la cantidad de especialidad y cursos a observar, y los datos existentes de ellos.

Por otro lado, es importante recordar que el último periodo de tiempo ha habido una externalidad importante que podría o no afectar los resultados, la cual corresponde a la pandemia, dado que no se sabe si esta pudo haber tenido alguna influencia en las motivaciones y postulaciones efectivas de los estudiantes, además del desempeño y vivencia de los actuales equipos docentes.

# Capítulo 2

## Marco de Referencia

### 2.1. Perspectiva de Género

De acuerdo con el texto “Perspectiva de Género” de Marta Lamas (Lamas,sf), el primer paso para establecer la perspectiva de género es comprender la diferencia entre sexo y género, en donde el primer término hace referencia a lo biológico, y el segundo término a aquello construido socialmente, es decir al conjunto de ideas, prescripciones y valoraciones sociales sobre lo masculino y lo femenino. A partir de lo anterior las culturas establecen un conjunto de prácticas, ideas, discursos y representaciones sociales que atribuyen como características específicas a mujeres y a hombres. Esta construcción simbólica que en las ciencias sociales se denomina género, reglamenta y condiciona la conducta objetiva y subjetiva de las personas, fabricando las ideas de lo que deben ser los hombres y las mujeres, de lo que se supone es ”propio” de cada sexo. Se ha comprobado que el estatus femenino varía de cultura en cultura, pero siempre con una constante: la subordinación política de las mujeres, a los hombres.

Por ello, según Lamas, se entiende por perspectiva de género, el desarrollo de una visión sobre los problemas de la relación hombre/mujer capaz de distinguir correctamente el origen cultural de éstos, comprender que la discriminación de las mujeres se produce de manera individual y colectiva, deliberada e inconsciente pues está tejida en las costumbres y la tradición, y se manifiesta en ataques directos a sus intereses o a ellas mismas y en ataques indirectos, provocados por el funcionamiento del sistema social o por la aplicación de medidas, de apariencia neutral, que repercuten especialmente en ellas debido a que se encuentran en peores condiciones para soportar sus efectos, o porque reúnen las condiciones para que se concentren en ellas los efectos perjudiciales de cierta actividad.

Dado lo anterior la autora recomienda que una buena administración pública debe impulsar e implementar medidas pro-activas, afirmativas, que detecten y corrijan los persistentes, sutiles y ocultos factores que ponen a las mujeres en desventaja frente a los hombres. Comprendiendo, además, que una perspectiva de género impacta a mujeres y a hombres, y beneficia al conjunto de la sociedad, al levantar obstáculos, discriminaciones y al establecer condiciones más equitativas para la participación en la sociedad.

## 2.2. Brechas de Género y su importancia transversal

Según la OCDE, en el año 2021 la ocupación femenina chilena se encuentra 20 puntos porcentuales debajo de la ocupación masculina. Esta diferencia se percibe tanto en el ingreso de la mujer al mercado laboral, como en la ocupación de empleos, el tipo de empleo y también la remuneración que emana de este, afectando negativamente al desarrollo pleno de la mujer laboralmente. En base a datos entregados el 2020 por el Ministerio de Desarrollo Social y Familia, las brechas salariales se mantienen en alrededor de 24%, sumado a la sobrerrepresentación que tienen en los sectores más productivos, como por ejemplo la minería. De esta manera, el problema se presenta multifactorialmente, por lo que su abordaje debe ser de la misma manera.

No se puede desconocer la evolución en la tasa de participación económica de las mujeres estos últimos casi 30 años. Según datos del INE, ha habido un incremento en más de un 10% en esta área; de un promedio de 41,6% de mujeres iguales o mayores a 15 años en 1990 a un 54% de mujeres en 2013, la tasa laboral femenina ha alcanzado positivos avances. Sin embargo, si se somete a un estudio comparativo, Chile sigue teniendo una de las tasas más bajas de toda la región (INE 2015). Por otro lado, y pese a la evolución observable en la tasa de participación económica, es importante tomar en cuenta los obstáculos simbólicos a los que se enfrentan las mujeres dentro del campo STEM, siendo estos factores importantes en la desigualdad de género. Según expresan Paula Szenkman y Estefanía Lotitto (2020), las mujeres en STEM comunican recibir una mayor exigencia que sus pares varones, tienen menos posibilidades de publicar artículos, experimentan una división tácita de roles según la cual se les asignan más tareas de cuidado, limpieza y secretaría, y se van afectadas por micromachismos dentro del campo laboral (9-10).

## 2.3. STEM y Sesgos

Como se mencionó anteriormente, el desarrollo de las disciplinas STEM son vistas como una posible respuesta a las distintas dimensiones que se dan en la problemática de género en el mercado laboral. La sigla STEM amalgama a cuatro disciplinas, claves en el desarrollo integrativo tanto a nivel social, económico, político e incluso cultural de la mujer. Las disciplinas en cuestión son Ciencia, Tecnología, Ingeniería y Matemáticas. El objetivo principal de STEM es potenciar estrategias de competitividad de los países en las esferas de innovación y desarrollo, “(...) así como también para responder a las necesidades de promover habilidades complejas que las personas requieren para desenvolverse en la sociedad actual, particularmente en el mundo del trabajo” (Jorquera 2020, 2). Es significativamente su nexo transversal con el sistema educativo lo que hace de STEM un sistema donde ahondar desde sus raíces la problemática de la perspectiva de género en la sociedad. El impacto social que busca promocionar STEM es la construcción de una sociedad inclusiva y sostenible, por lo que no se trata solamente del desarrollo de conocimiento, conductas y habilidades, sino también del cómo y con qué enfoque esto se promueve en la sociedad.

En ese sentido, el acceso a la educación es una variable crucial en STEM, ya que, como algunos expertos lo plantean, la brecha de género se asienta desde la primera infancia. Khan y



Ginther (en Cifuentes y Guerra 2020) expresan que la subrepresentación de la mujer comienza en la infancia, precisamente en el rendimiento en áreas claves de STEM como matemática en periodo escolar.

Existe una preconcepción y división patriarcal entre niños y niñas, aludiendo a diferencias cognitivas entre ambos géneros. El impacto de esta lógica es la generación, o, mejor dicho, la perpetuación de estereotipos arraigados culturalmente, como también reiteran Szenkman y Lotitto (2020): "tanto en los entornos familiares y escolares como en los medios de comunicación, existen estereotipos y sesgos de género que excluyen a las niñas y jóvenes de las ciencias e inciden en su menor participación en STEM. Incluso los juguetes publicitados con distinción de género se presentan como una influencia temprana en los intereses de niños y niñas, siendo estas últimas apartadas de las habilidades espaciales que se desarrollan a partir de juguetes tradicionalmente "masculinos". Esto, junto con la falta de visibilidad de mujeres que han destacado en áreas STEM, resulta en un bajo interés por el aprendizaje científico y matemático (8-9)."

Lo determinante de esto se manifiesta años después, en la elección académica y profesional de las mujeres. Como bien expresa Jorquera (2020), "A nivel internacional, las mujeres sólo son el 35 % de la matrícula estudiantil en áreas relacionadas con las ciencias, la tecnología, las matemáticas y la ingeniería" (4).

Se podría afirmar que existe una "ceguera de género", no sólo por parte de las instituciones ligadas a STEM, sino también por parte de las mismas mujeres. El punto acá no tiene que ver con responsabilizar a la misma mujer de no participar en cargos y empleos relacionados a la tecnología, matemática y ciencias, sino entender el por qué de la situación. Como se mencionó, la desventajas de mujeres y niñas en STEM comienza en una primera infancia, incrementando en el tiempo, visualizándose en la pérdida de interés que comienza particularmente en la adolescencia (Jorquera 2020).

Es aquí donde aparece uno de los sesgos más importantes, siendo el principal motivo por el que las niñas se alejan de STEM; este es la "autoselección". Según la UNESCO (2019), las mujeres se marginan de las áreas profesionales mencionadas a debido a diversos factores , los cuales son: factores personales/individuales, relacionadas a las habilidades lingüísticas, espaciales y numéricas, a la autoconfianza y motivación; factores familiares, relacionados a la promoción que tienen los mismos padres o miembros familiares para incentivar a las niñas a incursionar en áreas científico-matemáticas; factor escolar, cuyo vínculo debería promover un ambiente seguro e inclusivo, donde se reclutaran a profesores de ambos sexos y géneros, ampliando a su vez becas para estudiar e investigar; finalmente el factor social incide en la "autoselección", ya que se relaciona con la creación de políticas públicas y la legislación, sumado a la promoción de una imagen positiva y no sesgada de la mujer como profesional en STEM.

Se entiende entonces que existe la necesidad de actuar tanto dentro como fuera del campo STEM, de manera transversal e integradora. En otras palabras, debe existir voluntad política, tanto de las instituciones estatales como de la sociedad y su cultura en general. Según Glass y Minnotte (en Cifuentes y Guerra 2020), la presencia de más mujeres genera el rompimiento del círculo vicioso y plantea un círculo virtuoso donde se produzca mayor contratación de mujeres porque estas aplican más a los empleos y cargos.

Como bien expresa Cifuentes y Guerra, la problemática de la selección y autoselección se ha centrado en la oferta, causando un problema en el entendimiento de la cuestión en sí: el centrarse en la oferta reduce el problema a uno de carácter individual, no relacional, cuyo caso lo es. Por el lado de una investigación en la demanda, también se aprecia que los espacios laborales buscan reproducirse endogámicamente, obstruyendo el desarrollo pleno de las mujeres en los ambientes laborales/académicos (ibid. 5).

Se comprende que no hay razones biológica, sino ambientales que determinan la inserción de la mujer en STEM, que se manifiestan en la percepción y autopercepción de la mujer no sólo en el mercado laboral, sino también en la sociedad en general, inculcado por los estereotipos implícitos y explícitos de género (ibid.). Lo anterior conlleva a una merma en la entrada de la mujer a trabajos vinculados con STEM y en la permanencia en ellos. La deserción no tiene que ver con las condiciones de trabajo, sino con las variables que afectan el compromiso con el empleo, principalmente lo respectivo a la maternidad; sin embargo, se puede pensar que las condiciones relacionadas al equilibrio entre maternidad y trabajo debería ser un factor que los lugares de trabajo deban considerar, entregando mayor flexibilidad.

Según Glass (en Cifuentes y Guerra 2020), las pocas mujeres que ingresan a los trabajos en STEM, permanecen menos tiempo laburando que el resto de las mujeres que participan en otras prácticas profesionales; “(...) más de un 30 % de las mujeres de la población estudiada abandonan sus carreras en STEM, mientras que no más del 7 % lo hacen en otros trabajos profesionales” (Glass en Cifuentes y Guerra 2020, 6). Una de las causas puede ser la falta de incentivo tanto monetario como creativo; es principalmente el primero lo que se vincula con el concepto de tejado de vidrio o “glass ceiling”. Esta noción es para explicitar cómo las mujeres a pesar de tener un trabajo y una posición dentro del mercado laboral en STEM, están condicionadas a una relatividad en la permanencia de su trabajo. Por tal razón, junto a las ya mencionadas existe una penalización en el ingreso económico; penalización ya que la maternidad es incompatible con el mismo trabajo, por lo que se les tiende a pagar menos, incluso en casos cuando se incrementa los ingresos anuales, es generalmente para los hombres que para las mujeres (Aguirre, Matta y Montoya en Cifuentes y Guerra 2020).

El problema de la brecha de género en STEM se ve reforzado por lo que Paula Szenkman y Estefanía Lotitto denominan un círculo vicioso: puesto que las mujeres enfrentan mayores obstáculos que los varones para insertarse en el mercado de trabajo, acceder a empleos de calidad, sostener sus trayectorias laborales y ocupar puestos de decisión, consecuentemente tienen una menor presencia y visibilidad en áreas STEM. Esto ofrece menos modelos a seguir para las siguientes generaciones, lo cual reduce sus posibilidades e interés en dicho campo y refuerza sesgos de género existentes, perpetuando así el círculo vicioso. Si bien la investigación de Szenkman y Lotitto está enfocada en la política argentina, las soluciones que proponen las autoras para romper con la brecha de género existente podría iluminar caminos posibles para considerar en la presente investigación.

Para combatir el círculo vicioso ya mencionado y fortalecer la equidad de género en STEM, las autoras proponen la implementación de un plan integral que considere diversos frentes. En primer lugar, es de suma importancia reforzar la perspectiva de género en políticas socioeducativas. Para despertar mayor interés en las disciplinas STEM en niñas y jóvenes, y deconstruir la brecha de género que existe en aquellas disciplinas, se propone revisar los materiales y contenidos educativos para asegurar la visibilización de mujeres exitosas en las

áreas STEM, tradicionalmente masculinizadas.

Para contribuir a generar interés en aquellas áreas también se sugiere darle mayor importancia a la ciencia y tecnología en actividades extracurriculares, además de implementar medidas innovadoras y con perspectiva de género en orientación vocacional, para otorgar información acerca de las oportunidades laborales que las carreras en STEM ofrecen. Por otro lado, la realización de campañas con perspectiva de género que sensibilicen sobre la importancia de la equidad económica ayudaría a derribar sesgos de género y a promover la visibilidad de mujeres en STEM. Complementariamente, se incentiva la implementación de acciones afirmativas que promuevan la participación de mujeres en la ciencia y tecnología, como premios de reconocimiento y metas de género en posiciones de liderazgo (Szenkman y Lotitto 2020).

Finalmente, para facilitar la trayectoria e inserción de mujeres en STEM, se pueden establecer protocolos que impidan la desigualdad y violencia de género, tales como procesos de contratación libres de sesgos, el uso de lenguaje inclusivo en avisos de búsquedas laborales, la inclusión de un número mínimo de mujeres tanto en la preselección como en los comités de selección, la designación de equipos de trabajos diversos, y el diseño de evaluaciones basadas en habilidades. También se sugiere implementar programas de mentoría para identificar y fortalecer habilidades de liderazgo en mujeres, además de introducir beneficios diferenciales en los programas de promoción a áreas STEM para empresas que incorporen mujeres (Szenkman y Lotitto 2020).

## 2.4. Selección de cargos docentes en la Facultad de Ciencias Físicas y Matemáticas

La facultad se encuentra organizada en 12 departamentos, 5 de ciencias y 7 de ingeniería y en la actualidad cuenta con 5091 estudiantes de pregrado, 461 Académicos/as jornada completa y parcial y más de 1100 estudiantes de postgrado. (FCFM, sf)

Según cifras de la propia Escuela, al año 2019 contaba con 5472 estudiantes, con una matrícula femenina del 30 %, siendo los Departamentos de Ingeniería Industrial, Ingeniería Civil y Geología quienes reciben la mayor parte de estas estudiantes, con porcentajes superiores al 28 % en cada caso. En el caso contrario se encuentran los departamentos de Astronomía, Geofísica o Física, además de los departamentos de Ingeniería en minas, Matemática, Química e Biotecnología cuentan con no más del 5 %. En los últimos siete años se han graduado o titulado del orden de 564 profesionales por año, un 21 % de los cuales, en promedio, son mujeres. La tasa de retención promedio durante el primer año es de un 93 %, siendo las principales causas de abandono el retiro y eliminación por rendimiento académico. (FCFM, sf)

Como se mencionó anteriormente en el funcionamiento regular de la docencia de la facultad en general se encuentran tres agentes importantes, el profesor/a de cátedra, el estudiante y finalmente los auxiliares y ayudantes. Estas últimas figuras en general se encargan de asistir en términos de gestión y funcionamiento de cada curso, según le sea asignado por el profesor,

con actividades tales como docencia auxiliar, coordinación de ayudantes, coordinación de secciones paralelas de un curso, preparación y cuidado de equipo y material para las clases en el caso de los primeros, y por otro lado corrección de pruebas, ejercicios, tareas y laboratorios, asistencias y reclamos en el caso de los segundos. (FCFM, sf)

En el año 2016 entró en vigencia el documento “Normas generales sobre la participación de auxiliares y ayudantes en la docencia de pregrado de la FCFM” aprobado por el Consejo de Escuela, el cual detalla la existencia de dos cargos principales de apoyo a la Docencia en la FCFM: Profesores Auxiliares, quienes desempeñan las actividades que sean designadas por el profesor de cátedra o coordinador del curso como docencia auxiliar, coordinación de ayudantes, coordinación de secciones paralelas de un curso, preparación y cuidado del equipo y material para las clases en aula y en terreno, etc.; y Ayudantes, quienes desempeñan actividades como corrección de pruebas, ejercicios y tareas, ayudantes de laboratorio y salidas a terreno, etc.

Para la selección de profesores auxiliares o ayudantes cada semestre se debe abrir un concurso a través de una plataforma online U-Campus en la que cualquier estudiante que cumpla con los requisitos exigidos según curso y cargo podrá postular. La selección será resuelta por el profesor de cátedra, o en su defecto por el jefe Docente, Sub Jefe Docente, Jefe de Carrera o Coordinador del Curso según sea establecido por el Departamento respectivo. Ellos podrán observar en la misma plataforma la experiencia previa del estudiante como equipo docente y su nota de aprobación del ramo, además de su información personal.

Para ser nombrado en cargo los candidatos cumplir con:

1. Haber aprobado el curso o uno equivalente con una nota final superior a 5,0 o acreditar el conocimiento necesario en los contenidos.
2. Trayectoria académica destacada (promedio ponderado superior a 5,0), máximo de tres reprobaciones en los últimos dos años cursados y no tener inscripciones académicas extraordinarias aceptadas para el semestre en curso o los dos anteriores.
3. No haber sido sancionado por sumarios debidos a causas éticas o académicas durante su paso por la Facultad.
4. Firmar compromiso con el código ético para estudiantes de la FCFM
5. Preferencia a quienes acrediten realización y aprobación de cursos o talleres de capacitación en docencia.
6. En el caso de ayudantes de laboratorio, preferencia a quienes acrediten haber obtenido nota superior a 5,0 en los laboratorios del curso o equivalente

En la actualidad entonces, cualquier estudiante que cumpla con los requisitos antes mencionados puede ser seleccionado para algún cargo docente, pero no existen mayores lineamientos de los objetivos a nivel FCFM sobre lo que se quiere lograr con estos roles. Es decir, dentro de las actividades en la cuales un estudiante puede participar y adquirir mayor visibilidad, experiencias, contactos, entre otros beneficios, estos cargos son la mejor alternativa, por lo que es racional pensar que lo más óptimo para el desarrollo de los estudiantes sería participar alguna vez de esta experiencia, entre otras existentes en la universidad.

Entonces, dado que en la FCFM desde hace varios años esta buscando luchar contra las brechas de género y ser un organismo que promueve la participación de mujeres en las ciencias, es importante que se analice no solo el ingreso y permanencia de las estudiantes, sino que también su desarrollo, roles y busque posibles sesgos en otros espacios tan importantes, como en este caso los cargos docentes.

# Capítulo 3

## Diseño Metodológico

El diseño metodológico presentado responde a la pregunta de investigación planteada al utilizar una combinación de técnicas y enfoques que permiten realizar un análisis exhaustivo de los datos disponibles. El proceso se basa en la metodología de Descubrimiento de Conocimiento en Bases de Datos (KDD), que involucra la extracción de patrones y reglas a partir de los datos para su posterior análisis. Además, se emplea el Análisis Exploratorio de Datos (EDA) como una etapa inicial para comprender y visualizar la información. A continuación, se aplican modelos de aprendizaje de máquina supervisados, como Random Forest, Support Vector Machine (SVM), XGBoost y Regresión Logística, para clasificar las instancias en categorías predefinidas. Estas técnicas son complementadas con el uso de herramientas de evaluación, como la matriz de confusión y la validación cruzada, que permiten medir la calidad de las predicciones y evaluar el rendimiento de los modelos.

### 3.1. Descubrimiento de conocimiento en base de datos - Metodología KDD

Knowledge Discovery in Databases (KDD) es un proceso automático en el que se combina descubrimiento y análisis de una base de datos. En este proceso se extraen patrones, ya sea reglas o funciones, para posteriormente analizarlos, es decir se preparan los datos, se realiza minería de datos y finalmente se presentan los resultados. (Fayyad, Piatetsky-Shapiro Smyth, 1996)

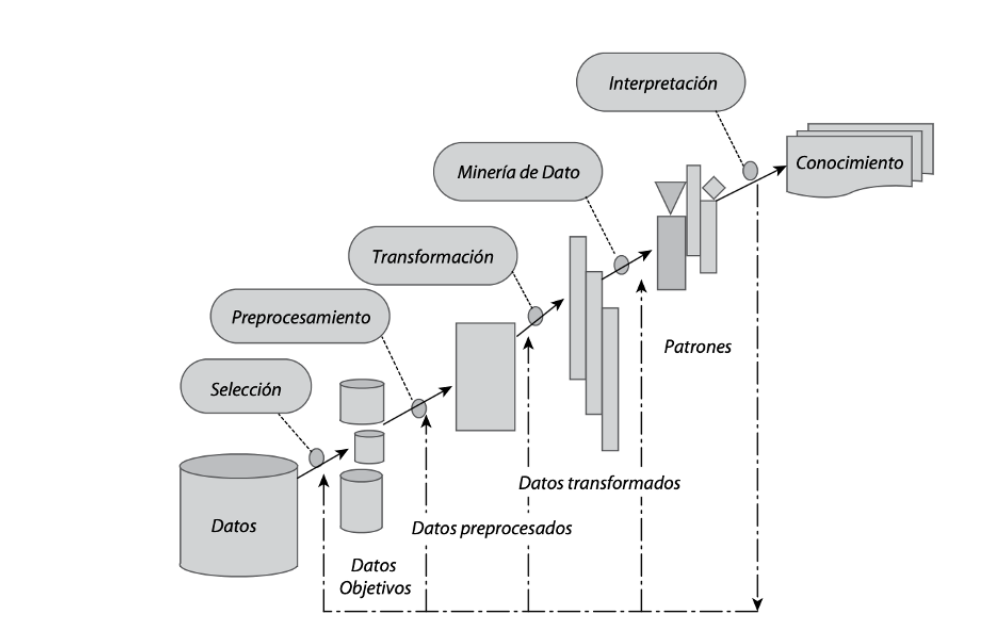
Si bien los pasos de esta metodología pueden variar, en este caso se realizarán los siguientes:

1. **Aprendizaje sobre el dominio del proyecto.** Se estudia el marco teórico relacionado con el tema para así recolectar conocimiento relevante para analizar los datos y construir los modelos con conocimiento y coherencia.
2. **Selección.** Una vez identificado el conocimiento relevante y prioritario y definidas las metas del proceso kdd, desde el punto de vista del usuario final, se crea un conjunto de

datos objetivo, seleccionando todo el conjunto de datos o una muestra representativa de este, sobre el cual se realiza el proceso de descubrimiento, esto utilizando el análisis exploratorio de datos.

3. **Preprocesamiento/limpieza.** Tratamiento de datos faltantes y outliers, en caso de aplicar. Se levanta información sobre los tipos de datos recolectados y se plantearan estrategias para abordar su estudio.
4. **Transformación/reducción.** Se buscan características útiles para representar los datos. Se utilizan métodos de reducción de dimensiones o de transformación para disminuir el número efectivo de variables bajo consideración o para encontrar representaciones invariantes de los datos.
5. **Minería de datos (data mining).** Aplicación de los modelos y algoritmos seleccionados, buscando identificar patrones.
6. **Interpretación/evaluación.** En la etapa de interpretación/evaluación, se interpretan los patrones descubiertos y posiblemente se retorna a las anteriores etapas para posteriores iteraciones. Esta etapa puede incluir la visualización de los patrones extraídos, la remoción de los patrones redundantes o irrelevantes y la traducción de los patrones útiles en términos que sean entendibles para el usuario. Por otra parte, se consolida el conocimiento descubierto para incorporarlo en otro sistema para posteriores acciones o, simplemente, para documentarlo y reportarlo a las partes interesadas; también para verificar y resolver conflictos potenciales con el conocimiento previamente descubierto.

Figura 3.1: El proceso KDD



Fuente: Communications of the ACM, 1996

## 3.2. Análisis exploratorio de datos (EDA)

El Análisis Exploratorio de Datos (EDA: Exploratory Data Analysis) fue definido por John W. Tukey en 1977 en su libro “Exploratory Data Analysis” como un tratamiento estadístico realizado previo al trabajo complejo con datos, con el fin de analizarlos y hacer el trabajo futuro más fácil y efectivo. (Tukey, 1997)

El EDA propuesto por Tukey no es una confirmación, sino que se basa en apariencias y búsqueda de insights a través de técnicas y métodos de aritmética simple y construcción de gráficos. Dentro de estos últimos se encuentran técnicas de conteo, redondeo, transformación numérica, resúmenes de variables utilizando media, mediana, ranking, construcción de gráficos para análisis visual como histogramas, boxplot, gráficos de tendencia lineales, punteados, paralelos, identificar valores anormales o ausentes, relación entre variables, comparación de variables, entre otras.

En base a la definición del libro “Mostly Harmless Econometrics” (RICE, 2006) se entiende el análisis multivariado de varianza (ANOVA) como una herramienta que se utiliza para comparar medias de dos o más grupos para una variable dependiente. El modelo estadístico se define como sigue:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

En donde  $Y_{ij}$  es la  $j$ -ésima observación del grupo  $i$ ,  $\mu$  es la media de todos los grupos,  $\alpha_i$  es la desviación media del grupo  $i$  y  $\varepsilon_{ij}$  representa otros factores que generan variación en el grupo. Lo que se debe testear es que la desviación media de los grupos es igual a cero.

En el presente estudio los análisis serán utilizados para observar y seleccionar las variables a utilizar en los modelos de clasificación.

## 3.3. Aprendizaje de máquina supervisados

En esta etapa se realiza la aplicación de la técnica de clasificación o predicción, para extraer conocimiento desde los datos. Se buscará a través de un modelo matemático categorizar cada instancia en las clases o grupos de estudio predefinidos.

Existen muchos modelos que pueden ser utilizados, ya sea para predecir números reales o en este caso categorizar la variable dependiente que es de tipo binomial o polinomial. En el caso de este estudio la variable es binomial y corresponde a si un estudiante fue o no seleccionado al cargo docente al cual postuló. Estos modelos o máquinas de aprendizaje buscarán predecir la decisión de los profesores con el fin de comprender qué variables afectan en este proceso. Las técnicas que serán utilizadas son Random Forest (RF), Support Vector Machine (SVM), XGBoost y Logistic Regression (LR).



### 3.3.1. Métodos de conjunto

Éstas son técnicas que crean múltiples modelos y luego los combinan para producir mejores resultados, por lo que en general producen soluciones más precisas que un solo modelo.

Para los métodos de conjunto de árboles de decisión se combinan varios árboles de decisión para producir un mejor rendimiento predictivo que utilizar un solo árbol. Además, se reduce uno de los principales problemas del árbol de decisión, que es el sobre ajuste. El principio principal detrás del modelo de conjunto es que muchos aprendizajes débiles se unen para formar así un aprendizaje fuerte. (González, 2021)

### 3.3.2. Random Forest

El método Random Forest, consiste en una gran cantidad de árboles de decisión individuales que operan como un conjunto, formando así un bosque. Dentro de este bosque, cada árbol individual genera una predicción de clase y la clase con más votos se convierte en la predicción del modelo.

### 3.3.3. Support Vector Machine

Este modelo es altamente utilizado por su buen comportamiento entre outliers, su objetivo es encontrar un hiperplano de separación que esté a igual distancia de los datos de entrenamiento más cercanos de cada clase. Así, se conseguirá el llamado margen máximo a cada lado del hiperplano. Además, cabe destacar que, para definir dicho hiperplano, solo se tendrán en cuenta, los datos que están justo encima del margen máximo. A estos puntos se les llama vectores soporte. (Valenzuela González, 2022)

### 3.3.4. XGBoost

Consiste en un ensamblado secuencial de árboles de decisión, en este caso los árboles se agregan secuencialmente con el fin de aprender del resultado de los árboles previos y corregir el error producido por los mismos, hasta que ya no se pueda corregir más dicho error (esto se conoce como “gradiente descendente”). (Espinoza, 2020)

### 3.3.5. Regresión logística

Este es un modelo que se suele utilizar para predecir variables binarias como es el caso de este estudio. Esta es una técnica estadística multivariante que permite estimar la relación existente entre una variable dependiente dicotómica y un conjunto de variables métricas y no métricas. El objetivo será encontrar las variables regresoras en la probabilidad de que un evento ocurra, en este caso que los estudiantes sean seleccionados. (Poma Mercedes, s.f.)

### 3.3.6. Herramientas de evaluación de los modelos

#### Matriz de Confusión

La matriz de confusión es una herramienta comúnmente utilizada en la evaluación de modelos de clasificación. Esta matriz permite visualizar y analizar la calidad de las predicciones realizadas por un modelo, mostrando la cantidad de casos que fueron clasificados correctamente y aquellos que fueron clasificados incorrectamente.

Una matriz de confusión típicamente muestra la cantidad de casos verdaderos positivos (TP), falsos positivos (FP), verdaderos negativos (TN) y falsos negativos (FN) en una tabla de 2x2. Un TP es un caso que fue clasificado correctamente como positivo, mientras que un FP es un caso que fue clasificado incorrectamente como positivo. De manera similar, un TN es un caso clasificado correctamente como negativo y un FN es un caso clasificado incorrectamente como negativo. Géron (2019)

#### Validación Cruzada

La validación cruzada es una técnica utilizada en el aprendizaje automático para evaluar el rendimiento de un modelo de manera más confiable y precisa. Consiste en dividir el conjunto de datos en  $k$  partes iguales, donde una de las partes se utiliza como conjunto de prueba y las restantes como conjunto de entrenamiento. Esto se repite  $k$  veces, cada vez con una parte diferente como conjunto de prueba, y se calcula la media de los resultados de las  $k$  iteraciones. (Hastie, Tibshirani, Friedman, 2009, p. 241)

## 3.4. Datos

El estudio cuenta con una base que integra los datos de la Subdirección de Gestión Docente, la cual se compone por 3296 datos estudiantes postulantes y seleccionados en cargos docentes de las carreras ingeniería en minas, matemáticas, eléctrica, industrial y geología entre los años 2017 y 2021. Dentro de estos datos se encuentran las variables género, fecha de nacimiento, curso al cual postula, carrera, cargo, estado de la postulación (Aceptada, pendiente o rechazada), semestre y año de postulación, nota con la que se aprobó el ramo, semestre, año en el que se cursó el ramo, género del profesor auxiliar que le realizó clases cuando tomó el curso, proporción de mujeres y hombres en la carrera cuándo postuló para el cargo y número de experiencias docentes anteriores al postular.

<b>Variable</b>	<b>Descripción</b>
Ids	Identificador único del estudiante
Sexo	Sexo del estudiante
Fecha_nacimiento	Fecha de nacimiento del estudiante
Edad	Edad actual del estudiante
EdadCargo	Edad al momento de postular al cargo
Carrera	Carrera del estudiante
Curso	Curso que postula el estudiante
Cargo	Cargo que postula el estudiante, puede ser Auxiliar o Ayudante
Estado_postulacion	Estado de la postulación del estudiante, puede ser Aceptado, Rechazado, Eliminado o en espera
AñoCargo	Año en que postuló al cargo
SemestreCargo	Semestre en que postuló al cargo
Nota_con_la_que_aprobo	Nota con la que aprobó el curso
Semestre_en_el_que_curso	Semestre en que cursó el curso
Añocurso	Año en que cursó el curso
CursosAnteriores	Cantidad de cursos previos en los que participó como equipo docente
Toma	Si postuló previo o post toma feminista
resultado	Si fue seleccionado o no
Proporción	Proporción de estudiantes del mismo género del estudiante en la carrera en el momento de postular
Genero_profesor	Género del profesor
Generos_iguales	Si el postulante tuvo un profesor auxiliar del mismo género cuando cursó el ramo

Tabla 3.1: Lista de variables y su descripción

# Capítulo 4

## Desarrollo metodológico

### 4.1. Análisis Exploratorio y Descriptivo de Datos

Para comenzar se realizó un análisis exploratorio y descriptivo de datos, con el fin de caracterizar a los estudiantes postulantes y también visualizar ciertos efectos que podrían tener ciertas variables en la postulación y selección.

#### 4.1.1. Equipos docentes vs proporción de género según carrera a través de los años

En primer lugar, parte de la hipótesis de esta investigación es que al aumentar el ingreso de las mujeres en los departamentos de la facultad, la representatividad de los cargos docentes también debiese aumentar. Para generar un análisis completo se realizó gráficos comparativos en donde se muestra la evolución según género de seleccionados en cargos docentes y la proporción según género por cada carrera.

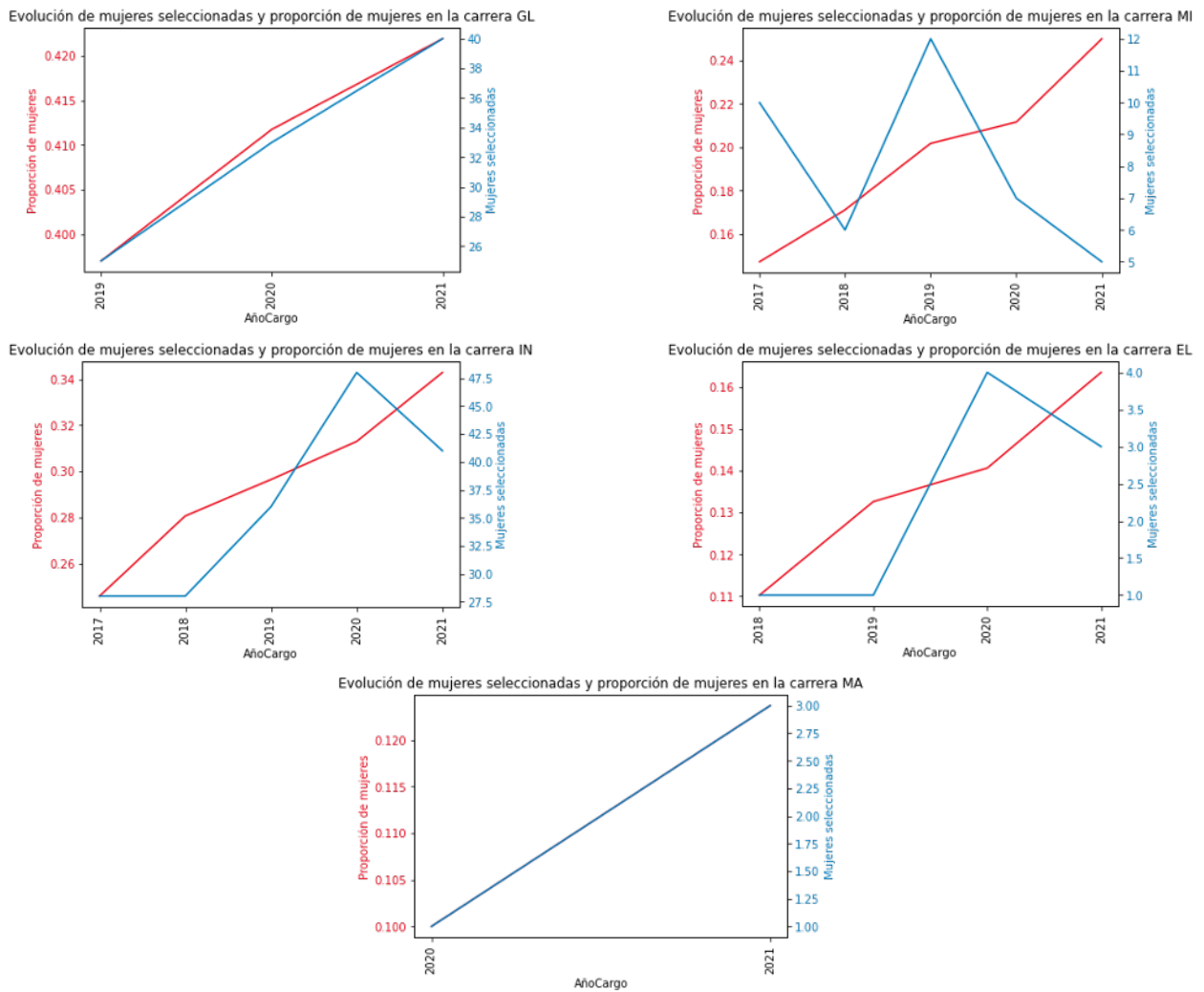


Figura 4.1: Evolución de mujeres seleccionadas y proporción de mujeres según carrera

Fuente: Elaboración propia

Al observar el comportamiento de las figuras 4.1 se puede notar que en todas las carreras la proporción de mujeres ha ido en ascenso, no así las seleccionadas en cargos docentes, que en su mayoría han tenido periodos de aumento y descenso notorios y lo más interesante es en el caso de las carreras eléctrica e industrias en las cuales ambas rectas se mueven de forma casi contraria. Por otro lado, es interesante que en el caso minas las mujeres seleccionadas después de un peak del año 2019, descendió notoriamente.

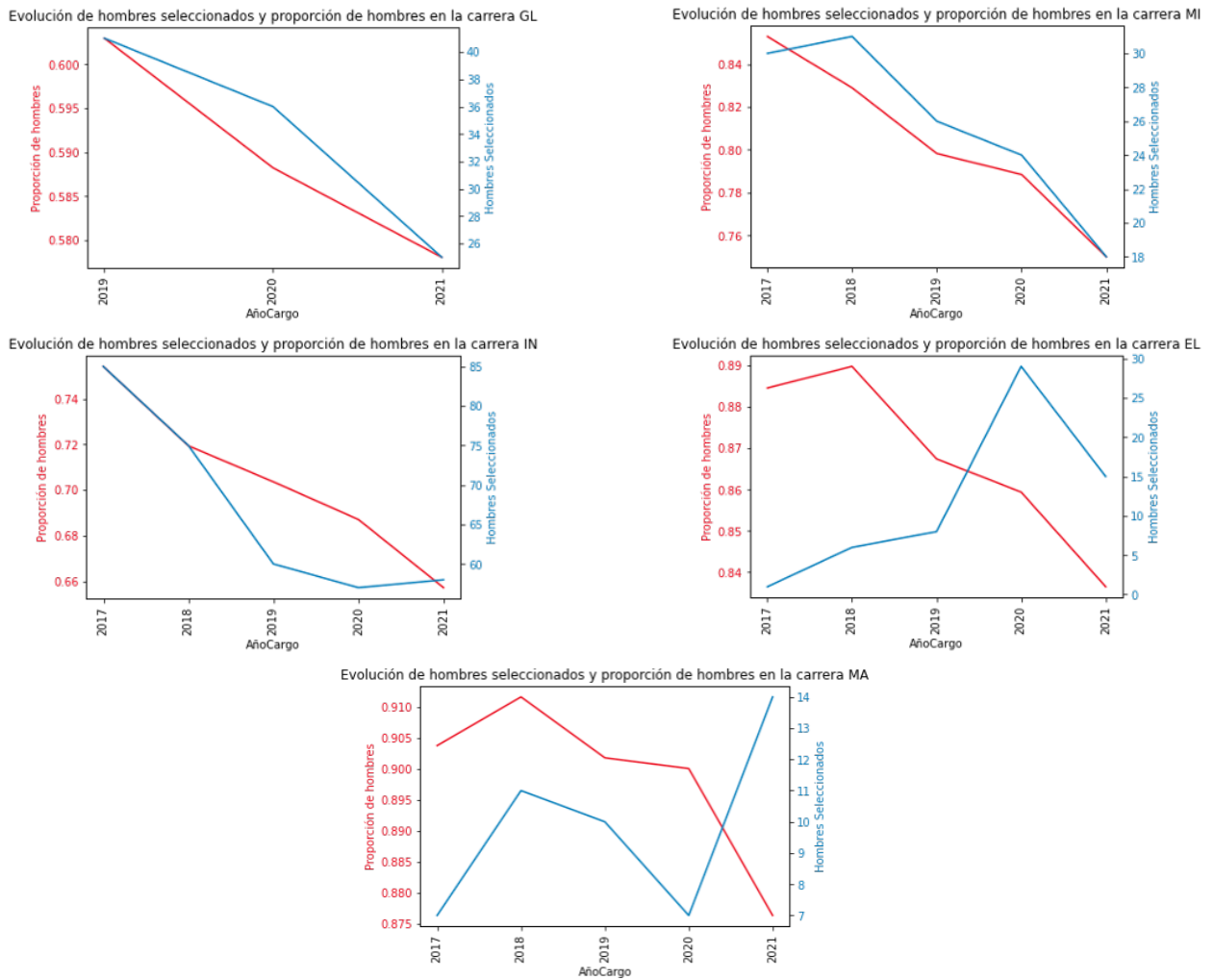


Figura 4.2: Evolución de hombres seleccionados y proporción de hombres según carrera

Fuente: Elaboración propia

Luego en el caso del otro sexo lógicamente la proporción de hombres a descendido, pero en general los seleccionados en el equipo docente también, exceptuando el caso de ingeniería en matemáticas.

#### 4.1.2. Postulantes vs proporción de género según carrera a través de los años

A partir del punto anterior es importante cuestionarse en los casos en que las mujeres y hombres seleccionados no siguen el patrón de ingreso de estudiantes a las carreras por género, en base a los sesgos estudiados en el marco conceptual es importante cuestionarse ¿Existe un sesgo de autoselección? Es por esto que, ahora se observa el comportamiento de las postulaciones.

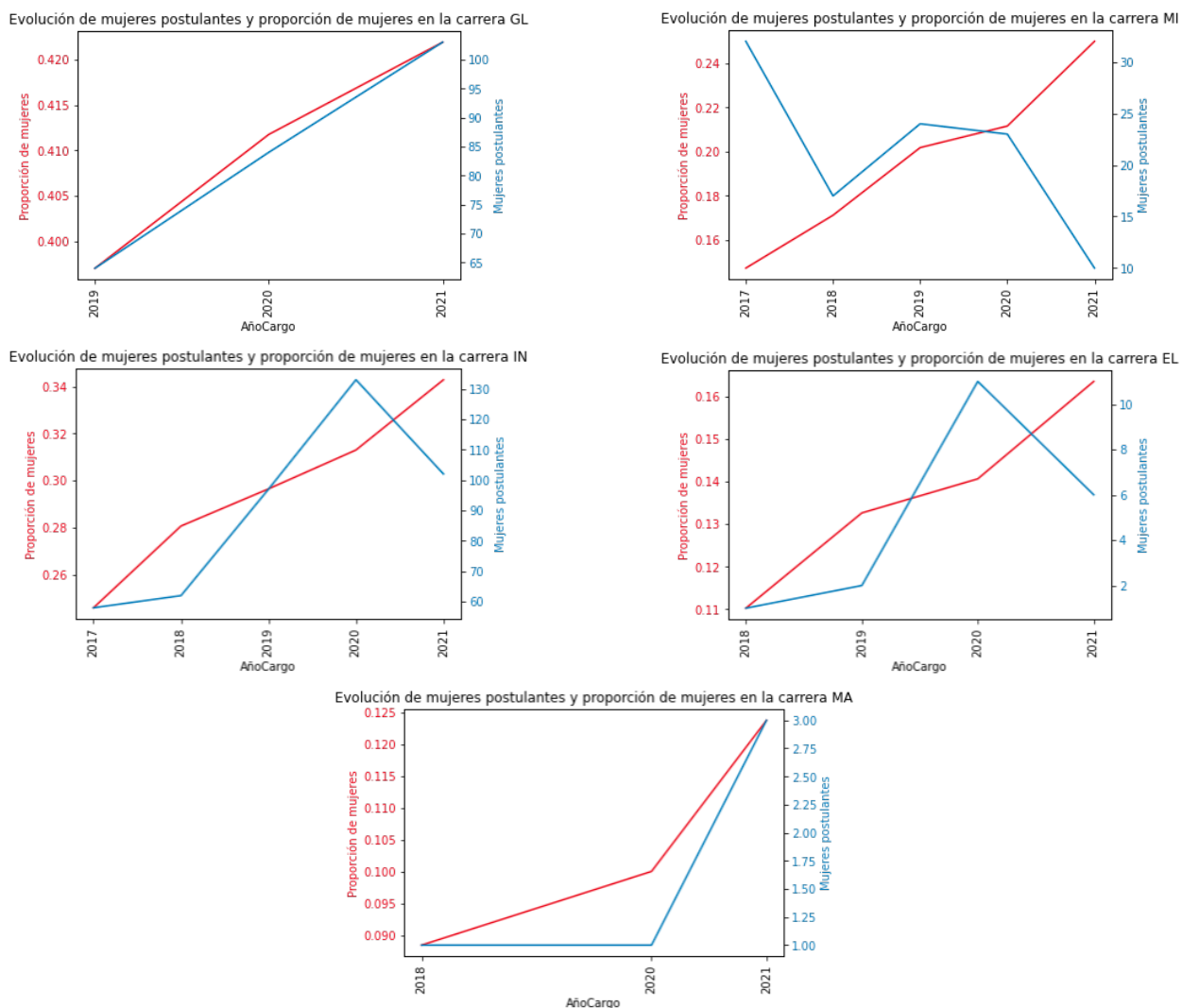


Figura 4.3: Evolución de mujeres postulantes y proporción de mujeres según carrera

Fuente: Elaboración propia

En el caso de las figuras 4.3 de las cinco carreras estudiadas minas, industrias y eléctrica tienen un comportamiento en su mayoría contrario en los últimos años al ingreso de mujeres a la carrera.

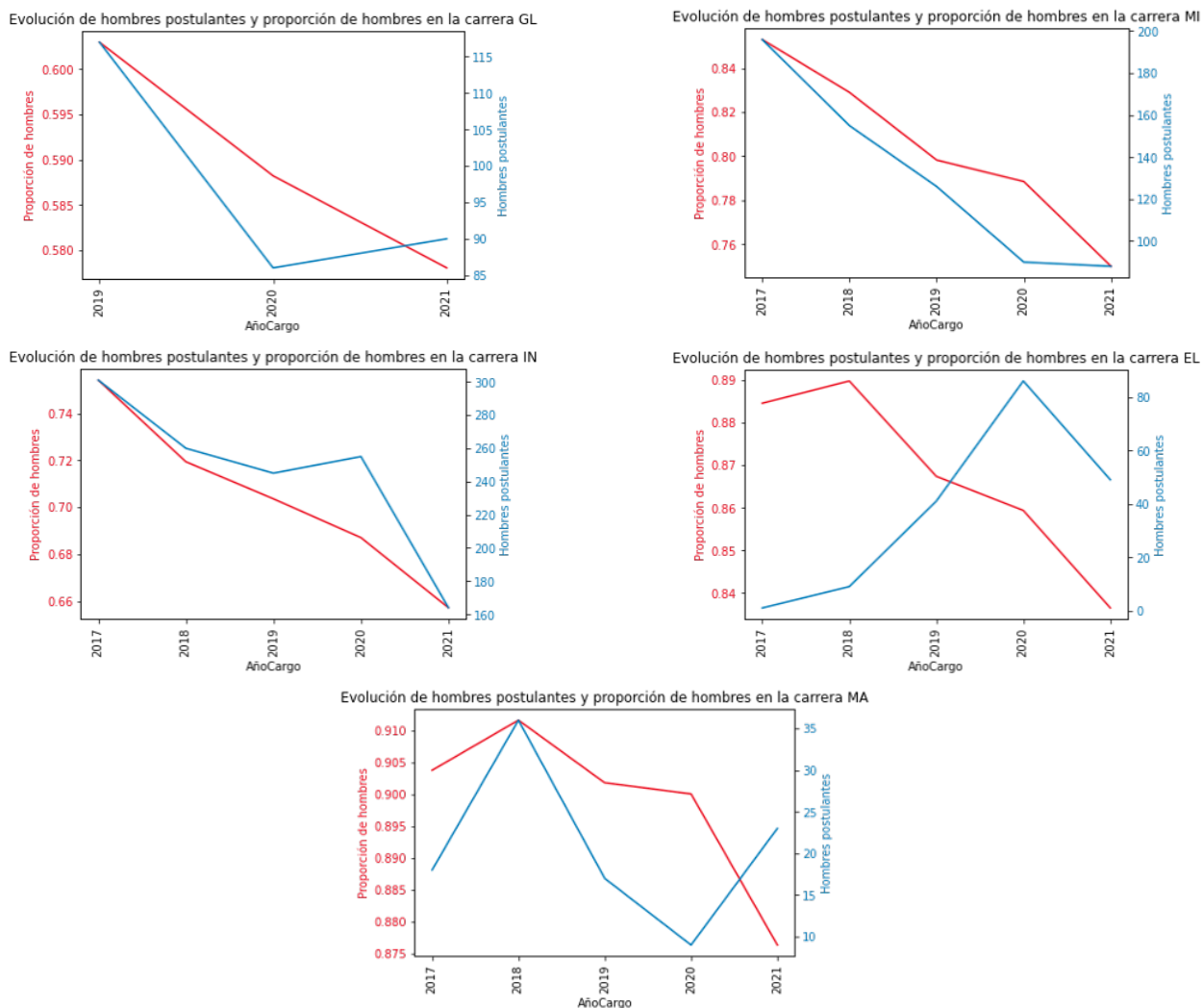


Figura 4.4: Evolución de hombres postulantes y proporción de hombres según carrera

Fuente: Elaboración propia

Ahora dos de las cinco carreras estudiadas, presenta un comportamiento contrario de las curvas en los últimos años. En ambos casos entonces se ve cierta relación entre postulantes y aceptados según género, que, si bien no son concluyentes, en cierta medida confirman que es necesario seguir indagando en los modelos con un enfoque de género

### 4.1.3. Proporción de postulantes vs Aceptados

Luego se decidió estudiar el comportamiento de las curvas de estudiantes postulantes y aceptados, considerando como hipótesis que las curvas debiesen moverse de forma similar en caso de no existir un sesgo.



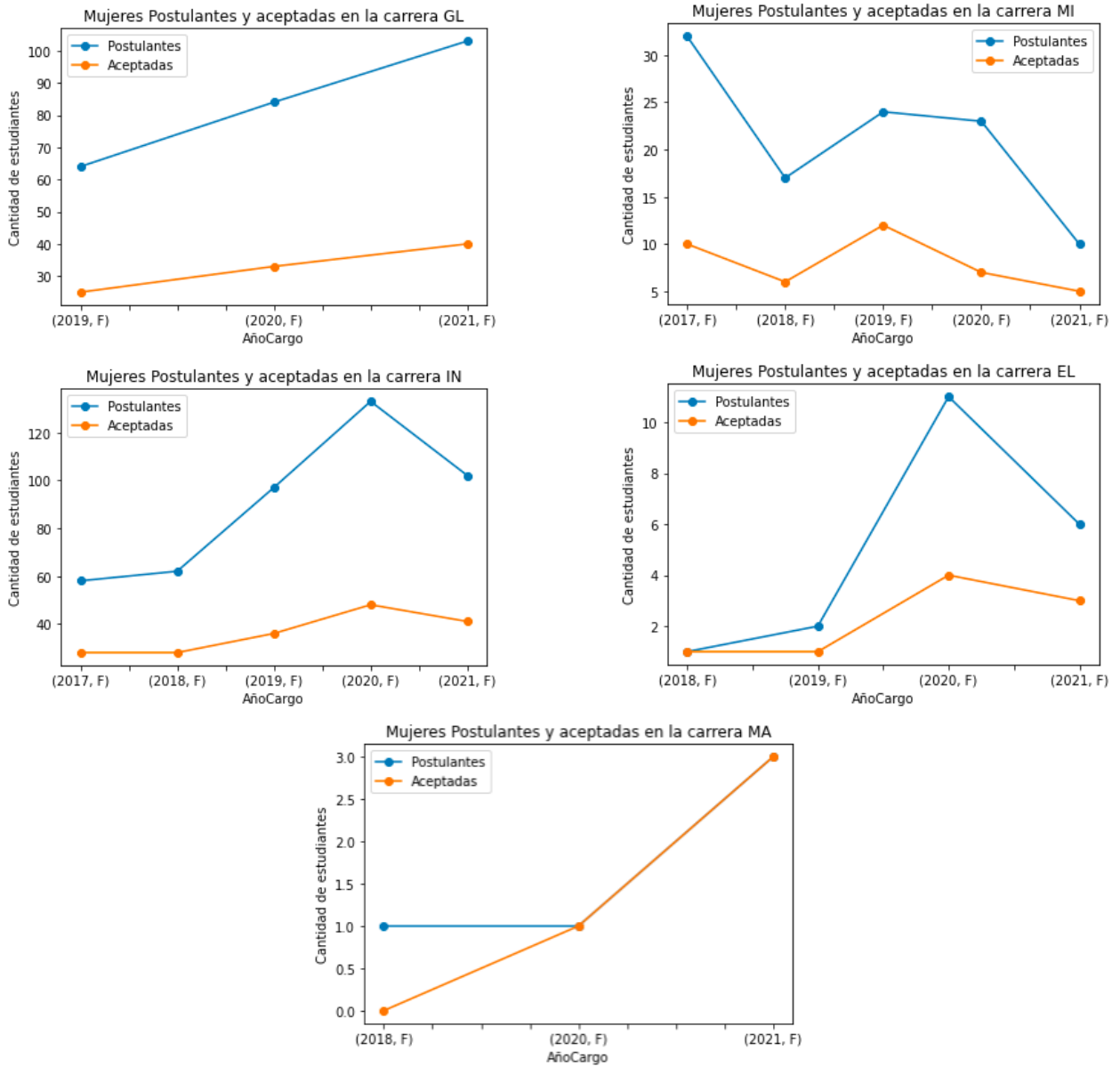


Figura 4.5: Evolución de mujeres postulantes vs aceptadas según carrera

Fuente: Elaboración propia

En el caso de las mujeres los comportamientos de postulantes y seleccionados son muy similares, lo que inicialmente mostraría que al parecer no existiría un sesgo de selección.

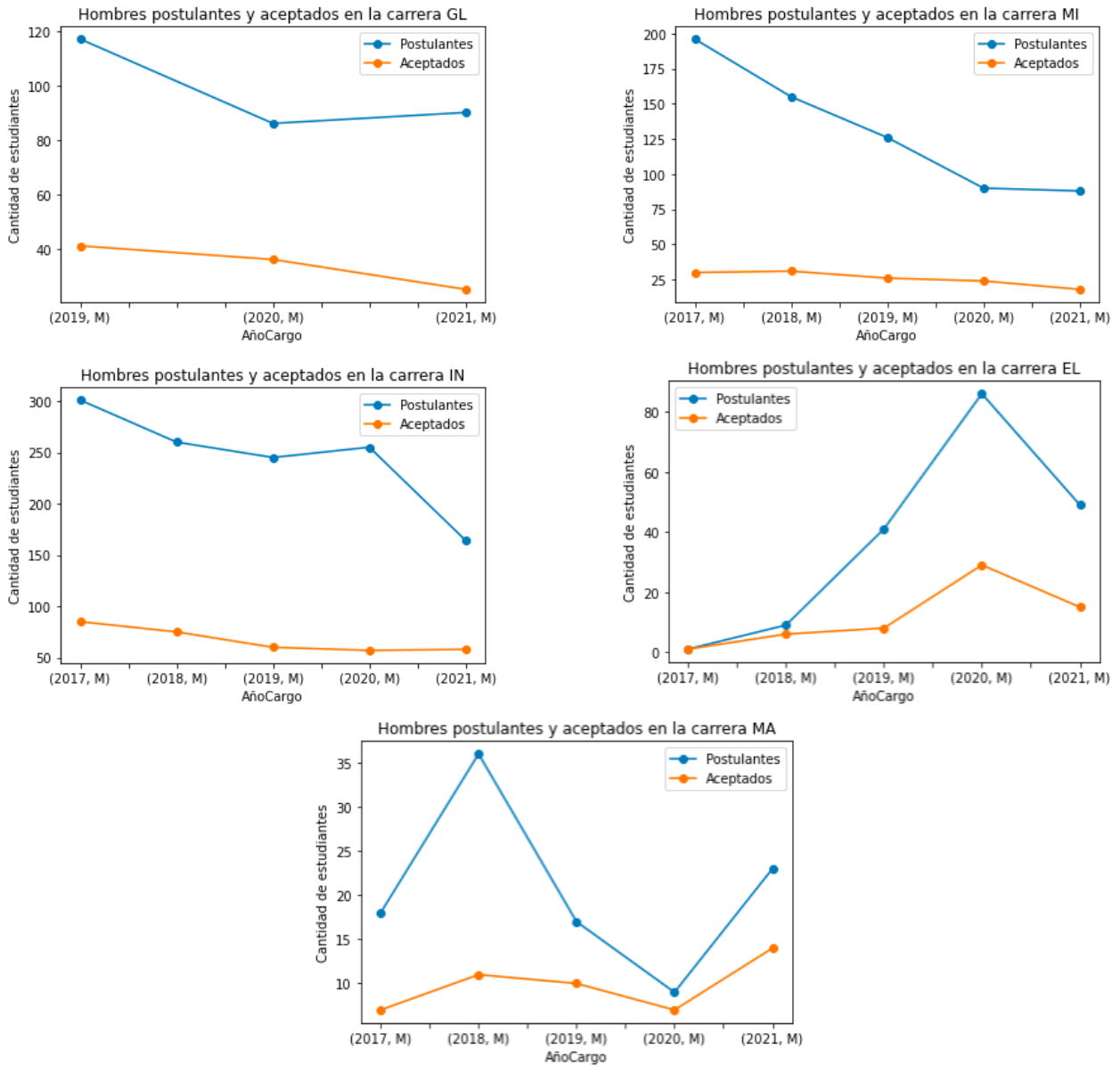


Figura 4.6: Evolución de hombres postulantes vs aceptados según carrera

Fuente: Elaboración propia

A diferencia de los gráficos antes estudiados (4.3) en este caso tanto para hombres como mujeres se ve un comportamiento similar entre curvas, lo que si bien no lo asegura daría un primer indicio de que no existe un sesgo de selección por género.

#### 4.1.4. Rendimiento académico del postulante según género

Una variable que debiese ser importante a la hora de seleccionar al estudiante es el rendimiento académico que este presentó al rendir el curso, dado que esto debiese demostrar el nivel de conocimiento que éste requiere para transmitir los contenidos del curso y realiza

revisiones de evaluaciones.

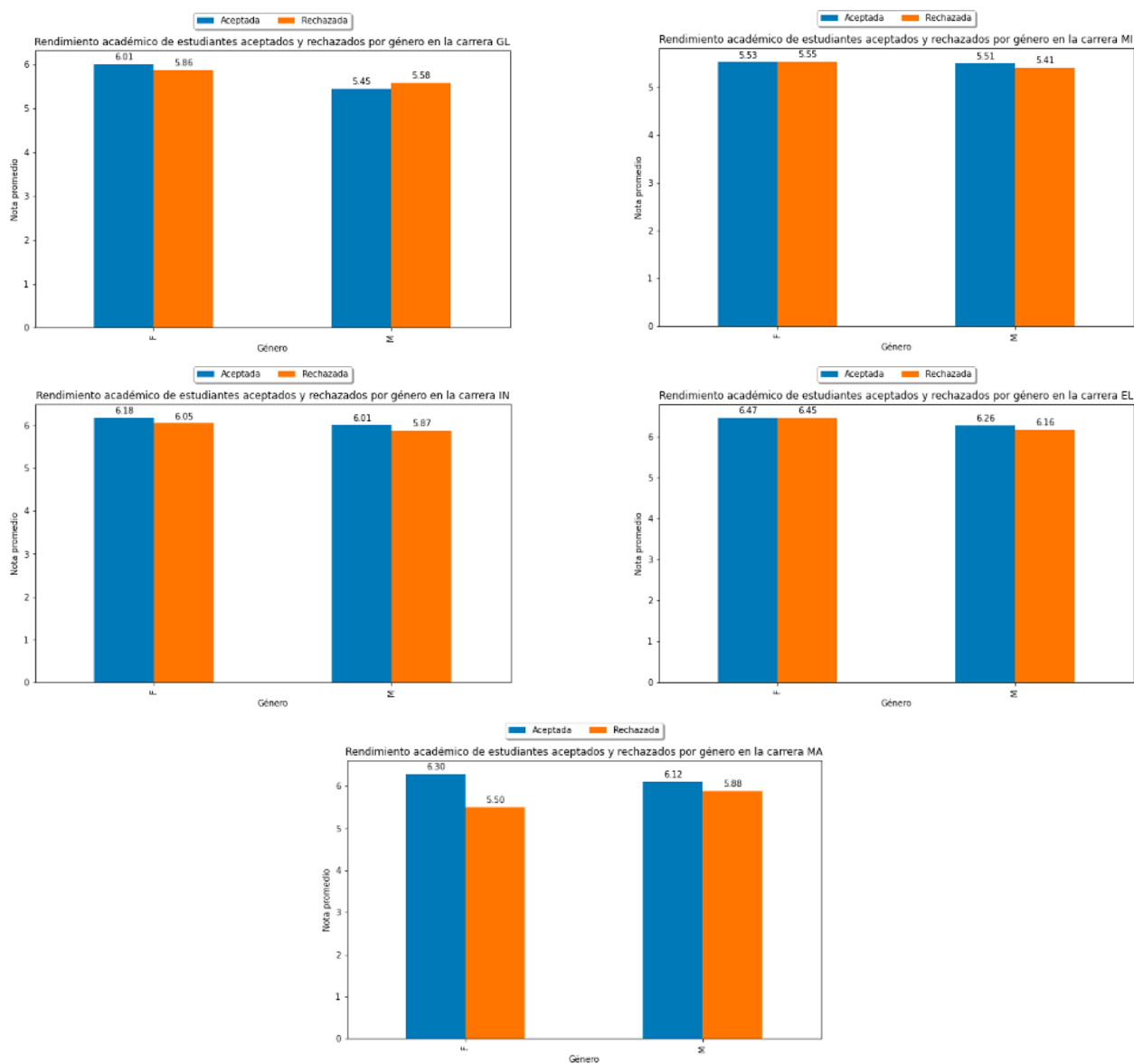


Figura 4.7: Rendimiento académico de estudiantes aceptados y rechazados por género para cada carrera

Fuente: Elaboración propia

En su mayoría se ve un comportamiento que parecería racional de parte de la selección, es decir los seleccionados tienen un mejor promedio de notas que los rechazados. Esto se cumple independiente del género exceptuando para los hombres de geología quienes presentan casi una décima en promedio de diferencia favorable para los rechazados. Por otro lado en todas las carreras se ve que la comparación por género del rendimiento académico esta un poco más inclinada hacia una mejor nota promedio exigida hacia las mujeres, sobre todo en matemáticas.

### 4.1.5. Experiencia académica según género

Otra variable importante para analizar es la experiencia como equipos docentes, en las que en general se espera que para el caso de hombres sea mayor dado que ya se observó en los gráficos 4.1 que históricamente las seleccionadas mujeres son menos que los hombres.

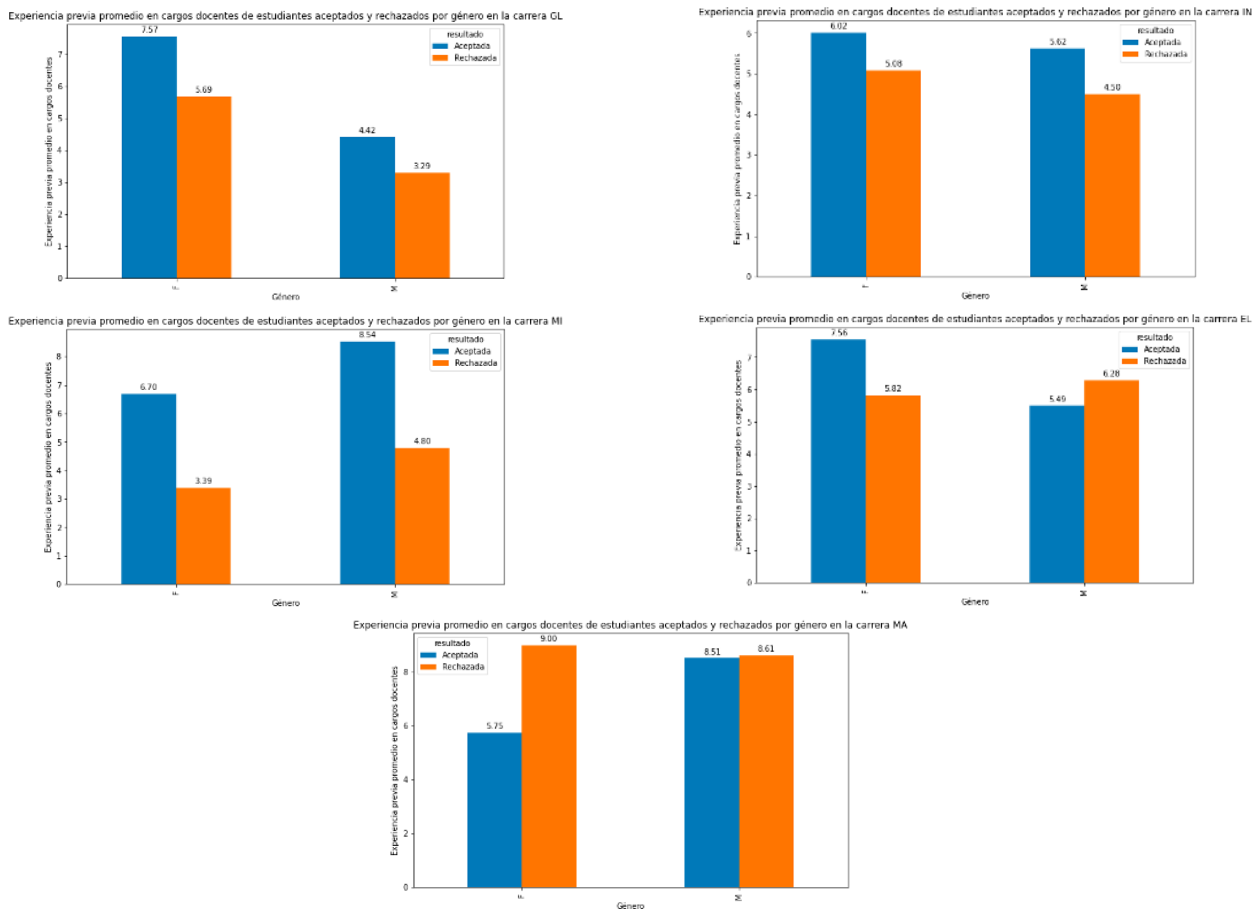


Figura 4.8: Experiencia previa promedio en cargos docentes de estudiantes aceptados y rechazados por género para cada carrera

Fuente: Elaboración propia

Como se puede observar en general todos los aceptados tenían más experiencia como equipos docentes, pero en el caso de los hombres en ingeniería eléctrica y las mujeres en matemáticas esto no se cumple. En el caso de la última carrera esto presenta lógica dado que históricamente habían muy pocas mujeres en los equipos docentes, si se observó un aumento más pronunciado de seleccionadas como se ve en los gráficos 4.1. estas estudiantes seleccionadas no debieron presentar gran experiencia.

### 4.1.6. Experiencia académica según género

Como se menciona en el marco conceptual los modelos de rol son importantes para aumentar la participación de las mujeres en la STEM, en este caso es interesante entonces observar cuántas estudiantes efectivamente tuvieron una mujer como auxiliar al cursar el ramo al cual después postuló. Es importante considerar que de esta variable se escapan todos los otros cursos que la estudiante pudo haber rendido en los que podría haber existido un modelo de rol para la estudiante.

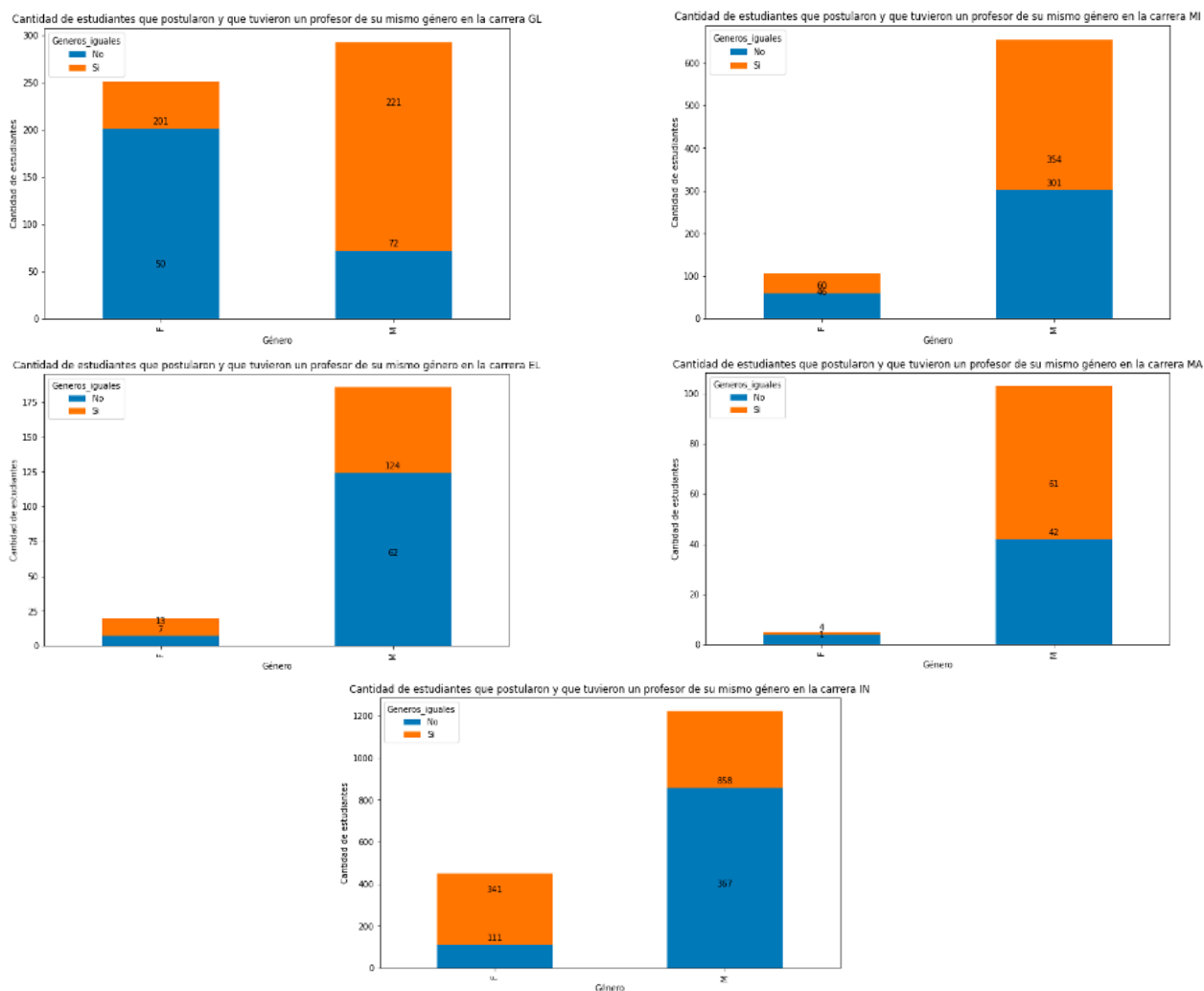


Figura 4.9: Cantidad de postulantes que tuvieron un profesor auxiliar de su mismo género al cursar el ramo

Fuente: Elaboración propia

En este caso, tanto para hombres como mujeres se cumple que la mayoría de las estudiantes aceptados tuvo un profesor auxiliar de su mismo género, lo que refuerza la creencia de que efectivamente esto tiene un efecto en la postulación.

### 4.1.7. Gráficos de distribución

En primer lugar, se decide analizar la distribución de edades de los postulantes y aceptados por género con el fin de caracterizar estos datos un poco más.

Distribución de la edad al momento de postular para estudiantes aceptados

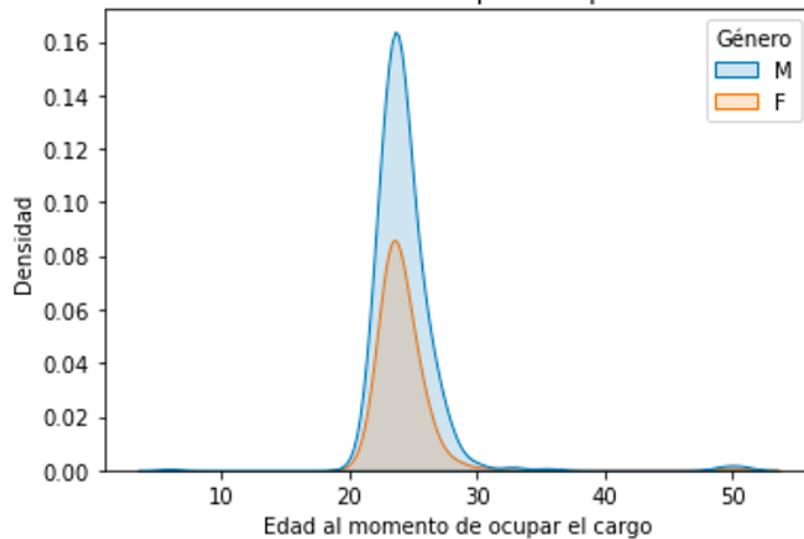


Figura 4.10: Distribución de la edad al momento de postular para aceptados

Fuente: Elaboración propia

Distribución de edad al postular por género

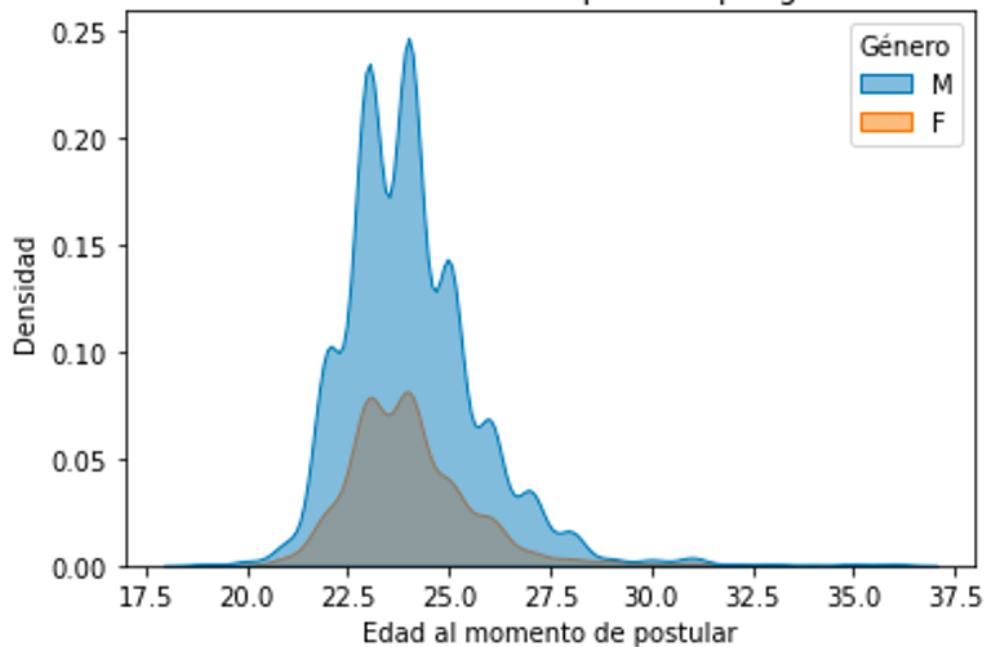


Figura 4.11: Distribución de edad al postular por género

Fuente: Elaboración propia

Al observar las gráficas se puede ver que los comportamientos de densidad son bastante similares para ambos géneros, es decir tanto para hombres como mujeres se prefieren estudiantes del mismo rango de edad a la hora de seleccionar.

Distribución de la nota con la que aprobaron el curso para estudiantes aceptados

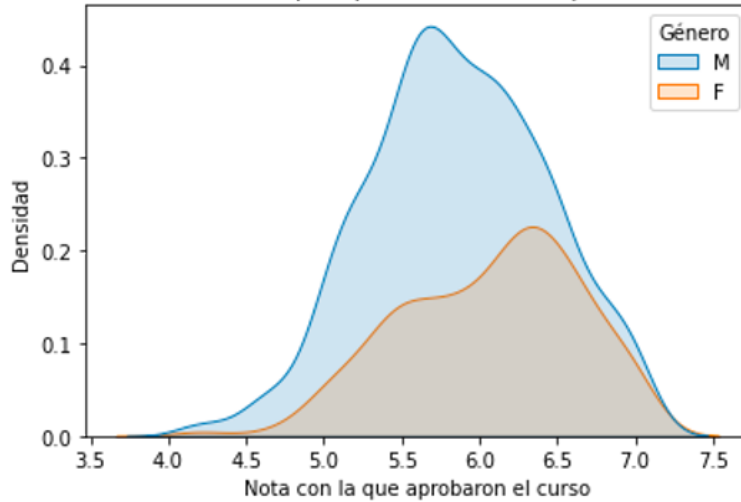


Figura 4.12: Distribución de la nota de aprobación para estudiantes aceptados por género

Fuente: Elaboración propia

Distribución de la cantidad de cursos previos participando como ED para estudiantes aceptados

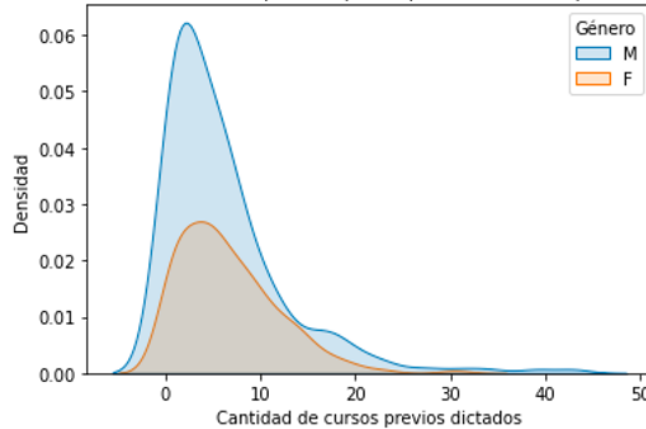


Figura 4.13: Distribución de experiencia como ED para estudiantes aceptados

Fuente: Elaboración propia

Al observar las distribuciones de las notas por género, podemos notar que en el caso de las mujeres aceptadas estas si presentan una mayor exigencia académica a la hora de ser seleccionadas como se había inferido anteriormente en 4.1.3, por otro lado, la cantidad de cursos previos dictados presenta un comportamiento similar para ambos géneros.

Ahora se realizaron gráficos de distribución separador por aceptados y rechazados para

cada variable no binaria o categórica, con fin de analizar su comportamiento para los modelos de predicción.

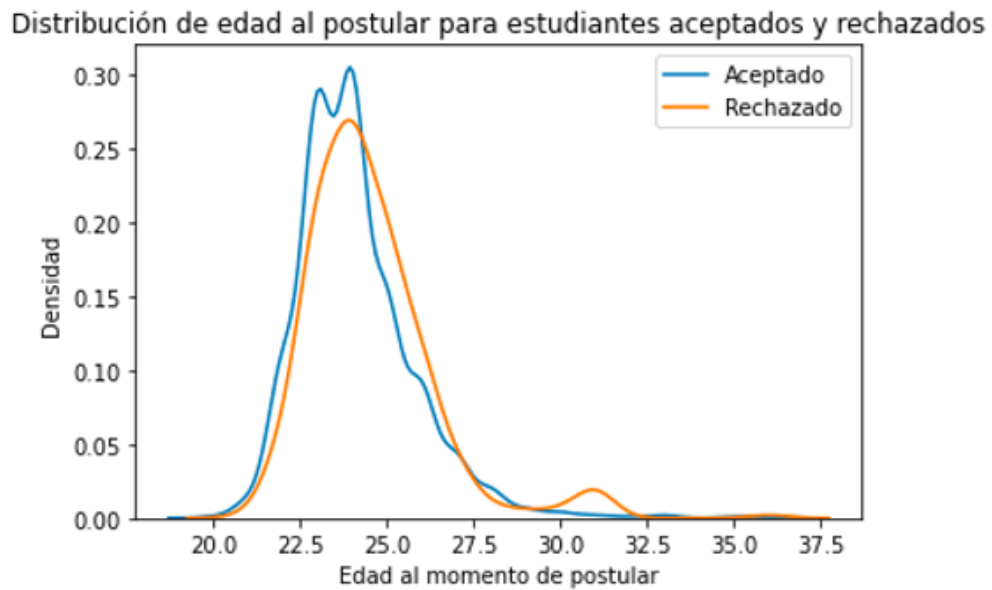


Figura 4.14: Distribución de edad al postular para estudiantes aceptados y rechazados

Fuente: Elaboración propia

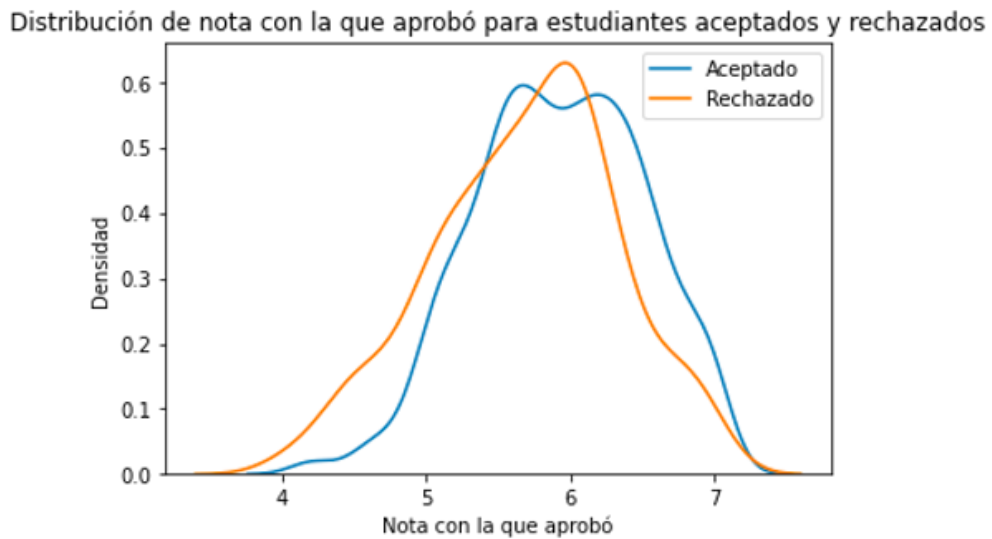


Figura 4.15: : Distribución de rendimiento académico para estudiantes aceptados y rechazados

Fuente: Elaboración propia



Distribución de cursos anteriores para estudiantes aceptados y rechazados

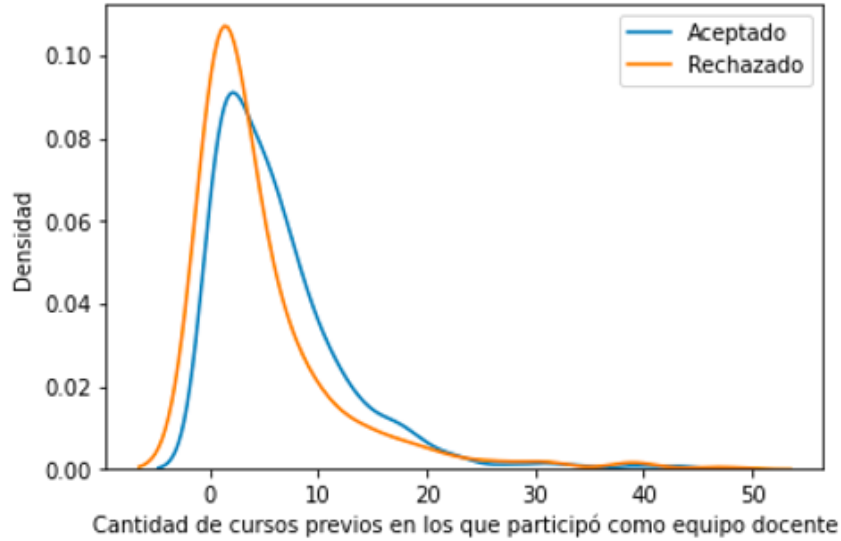


Figura 4.16: Distribución de experiencia como ED para estudiantes aceptados y rechazados

Fuente: Elaboración propia

Distribución de proporción de estudiantes del mismo género para estudiantes aceptados y rechazados

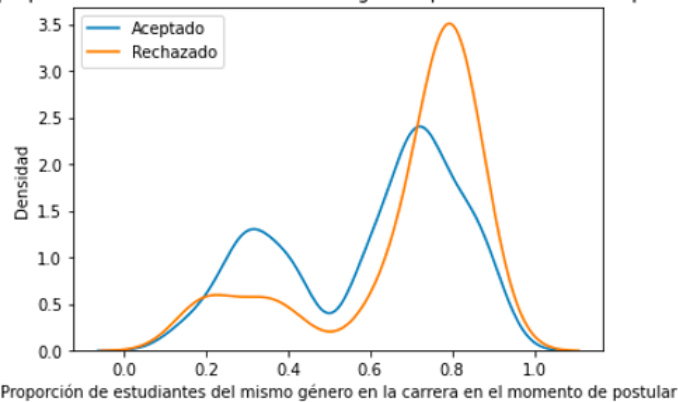


Figura 4.17: Distribución de proporción de estudiantes del mismo género para estudiantes aceptados y rechazados

Fuente: Elaboración propia

Estos gráficos permiten comprobar rápidamente que los datos presentan un comportamiento similar para ambos grupos, esto ayudaría a descartar relaciones no lineales entre las variables independientes y candidatos aceptados.

#### 4.1.8. Matriz de Correlación

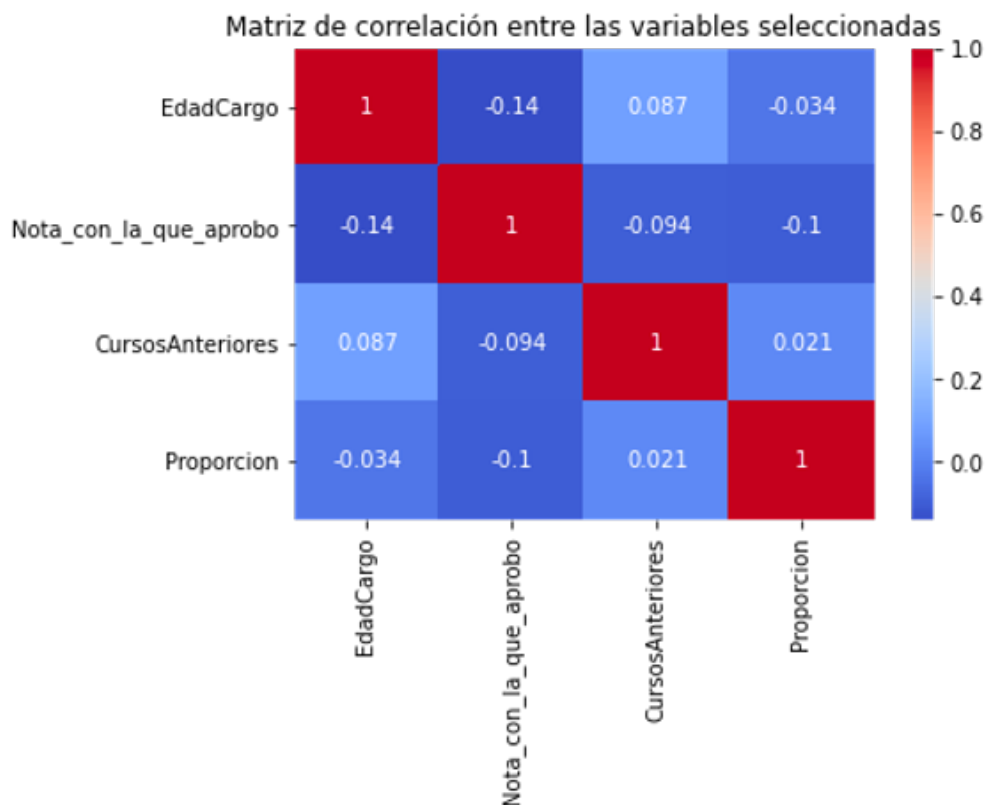


Figura 4.18: Mapa de calor de la matriz de correlación entre variables numéricas

Fuente: Elaboración propia

La matriz de correlación muestra la relación lineal entre las variables edad al postular al cargo, nota con la que aprobó, cursos anteriores como equipo docente y la proporción de estudiantes en la carrera a la hora de postular al cargo, todo esto para los estudiantes aceptados

Se puede observar que la variable "Nota con la que aprobo" tiene una correlación negativa moderada con "EdadCargo", lo que sugiere que a medida que la edad de los postulantes aumenta, su nota con la que aprobaron disminuye en promedio. Además, "Nota con la que aprobo" tiene una correlación negativa débil con "Proporción", lo que sugiere que a medida que la proporción de estudiantes en la carrera aumenta, la nota con la que aprobaron los postulantes disminuye en promedio.

La variable "CursosAnteriores" tiene una correlación positiva débil con "EdadCargo", lo que sugiere que a medida que la edad de los postulantes aumenta, es más probable que hayan tenido más cargos docentes anteriores en la carrera. También se observa una correlación positiva débil entre "CursosAnteriores" y Proporción, lo que sugiere que a medida que la proporción de estudiantes en la carrera aumenta, los postulantes tienden a haber tenido más cursos anteriores.

Finalmente, la variable "Proporcion" no tiene una correlación significativa con ninguna de las otras variables.

En general, la matriz de correlación sugiere que hay algunas relaciones interesantes entre las variables, pero también hay algunas correlaciones muy débiles o nulas que no permiten seleccionar una sobre la otra.

#### 4.1.9. ANOVA

Tabla 4.1: Resultados de la ANOVA para las variables de interés. Fuente: elaboración propia

Variable	F-value	p-value
EdadCargo	0.545137	0.460484
Nota_con_la_que_aprobo	29.2402	7.99117e-08
CursosAnteriores	0.879114	0.34867
Proporcion	5007.1	0

De la Tabla 4.1 se obtiene que en cuanto a la variable "Nota\_con\_la\_que\_aprobo", se encontró un F-value de 29.2402 y un p-value muy bajo ( $7,99117 \times 10^{-8}$ ). Estos resultados sugieren que hay diferencias significativas en las notas con las que los estudiantes aprobaron, lo que puede ser un factor relevante en la selección de los equipos docentes.

Por otro lado, la variable "Proporcion" muestra un F-value muy alto de 5007.1 y un p-value de 0. Esto sugiere que existen diferencias significativas en la proporción de mujeres y hombres en la carrera al momento de postular para los cargos docentes.

En el caso de las otras variables no se puede concluir que tengan un efecto significativo en la variable dependiente debido a sus p-values mayores a 0.05.

#### 4.1.10. Test de hipótesis

Para seguir analizando si existe algún tipo de sesgo en la selección de estudiantes por su género se decidió realizar un test de hipótesis utilizando la función chi2.

La tabla de contingencia que indicó que de un total de 2462 postulantes masculinos, 674 (27.4%) fueron aceptados para el cargo al que se postularon, mientras que de un total de 834 postulantes femeninos, 332 (39.8%) fueron aceptados. En términos absolutos, se aceptaron más hombres que mujeres (674 vs 332), pero esto no necesariamente indica una relación significativa entre el género y la selección para el cargo.

La prueba de hipótesis arrojó un valor de chi-cuadrado de 45.40 y un p-valor de 3.28e-09. Esto indica que hay una relación significativa entre el género del postulante y su selección para el cargo, ya que el p-valor es mucho menor que el nivel de significancia de 0.05.

En otras palabras, la hipótesis nula de que no hay relación entre el género del postulante y su selección para el cargo se rechaza, lo que sugiere que el género del postulante es un

factor que influye en la selección para el cargo. Es importante tener en cuenta que la relación entre el género y la selección para el cargo puede ser influenciada por otras variables que no se consideraron en este test de hipótesis, es por ello que se debe continuar analizando.

Tabla 4.2: Tabla de contingencia y resultados del test de hipótesis. Fuente: Elaboración propia

	M	F	Total	
0	1788	502	2290	Chi-cuadrado: 45.4038 p-valor < 0.001
1	674	332	1006	
All	2462	834	3296	

# Capítulo 5

## Desarrollo de los modelos

Como ya se ha mencionado, el presente trabajo de título busca encontrar los principales factores que influyen en la selección de un estudiante a un cargo docente. Dado lo anterior, y sumado a las limitaciones causadas por la naturaleza de los datos utilizados, se proponen modelos simples, aplicados por la literatura en problemas similares. Los modelos seleccionados son Random Forest, modelo XGBoost y Regresión Logística, todos adaptados a problemas de clasificación.

### 5.1. Definición de variables

Para el presente estudio, se decidió utilizar las variables que se pueden ver en la Tabla 5.1.

Tabla 5.1: Descripción de las variables. Fuente: Elaboración Propia

Variable	Descripción
EdadCargo	Edad al momento de postular al cargo
SemestreCargo	Semestre en que se postuló al cargo
Nota_con_la_que_aprobo	Nota con la que aprobó el curso
CursosAnteriores	Cantidad de cursos previos en los que se participó como equipo docente
Proporcion	Proporción del género del estudiante en la carrera en el momento de postular
Cargo_3	Cargo Auxiliar
Cargo_2	Cargo Ayudante
Género_F	Género Femenino
Género_M	Género Masculino
Carrera_GL	Carrera Geología
Carrera_IN	Carrera Ingeniería civil industrial
Carrera_MA	Carrera Ingeniería civil matemáticas
Carrera_MI	Carrera Ingeniería civil minería
Toma_Pretoma	Si la postulación fue realizada antes de la toma feminista o después
Genero_profesor_F	Profesor auxiliar cuando el estudiante realizó el curso fue mujer
Generos_iguales_Si	Si el postulante tuvo un profesor del mismo género

Esto se basa principalmente en que los resultados del análisis exploratorio no permitieron descartar variables debido al bajo impacto de cada una sobre la selección, pero por sobre todo dada la limitada cantidad de variables existentes en la base, esto debido a que para ciertas carreras la participación de mujeres en equipos docentes en algunos años no existió o no fue registrado en U-Campus, esto hizo que exista un bajo costo computacional del entrenamiento de cada modelo y se optó por no implementar métodos automáticos de selección de variables en esa iteración.

## 5.2. Estructura y optimización de los modelos

Para la construcción de los modelos se necesito hacer una balanceo de las clases y aplicación del algoritmo de clasificación. Debido a que existía una gran diferencia entre tamaños de aceptados y rechazados se utilizó un algoritmo de balanceo utilizando SMOTE.

Este algoritmo aumenta la cantidad de ejemplos de la población estudiada generando ejemplos sintéticos basados en la interpolación de datos existentes. Esto ayuda a mitigar el desequilibrio de clases y proporciona un conjunto de datos más equilibrado que se puede utilizar para entrenar modelos de aprendizaje automático de manera más efectiva.

Posteriormente los hiperparámetros de cada modelo fueron optimizados por medio de

GridSearchCV, una clase que recibe una lista de valores para cada parámetro de un modelo, para luego elegir empíricamente la combinación óptima de hiperparámetros.

### 5.3. Entrenamiento y resultados

El resultado del entrenamiento de cada modelo se obtuvo en base a una validación cruzada Stratified-K-fol. Las métricas elegidas para comparar rendimientos entre los modelos fueron recall y weighted recall. La muestra de entrenamiento se divide en conjuntos de entrenamiento y prueba utilizando la función `train_test_split` con un tamaño de prueba del 30 % de los datos totales. Por otro lado el umbral de predicción utilizado en los modelos es de 0.5, esto dado que este es un problema de clasificación binaria, es de los estudiantes o son seleccionados o no lo son.

El recall es una medida de la capacidad del modelo para identificar las instancias positivas, los resultados de cada modelo fueron los siguientes:

- Para el modelo SVM, el recall ponderado es de 62.08 % lo que indica que el modelo es capaz de identificar correctamente el 61.28 % de las instancias positivas en el conjunto de prueba.
- El modelo XGBoost obtiene un mejor rendimiento con un recall ponderado del 74.6 % y un recall del 74.89 %. Esto sugiere que el modelo es mejor para identificar correctamente las instancias positivas.
- Para el modelo de Random Forest, el recall ponderado es del 70.62 % y el recall del 70.60 %, lo que indica que el modelo tiene un rendimiento moderado en términos de identificación de instancias positivas.
- Finalmente el modelo de regresión logística tiene el peor rendimiento en término de recall ponderado (63,25 %) y recall (63.46 %).

Para complementar esta información se decidió hacer una matriz de confusión por modelo, las cuales son las siguientes:

Tabla 5.2: Matriz de confusión SVM. Fuente: Elaboración propia.

	Predicción negativa	Predicción positiva
Verdadero negativo	521	166
Falso negativo	366	321

Tabla 5.4: Matriz de confusión Random Forest. Fuente: Elaboración propia.

	Predicción negativa	Predicción positiva
Verdadero negativo	488	199
Falso negativo	205	482

Tabla 5.3: Matriz de confusión XGB. Fuente: Elaboración propia.

	Predicción negativa	Predicción positiva
Verdadero negativo	556	131
Falso negativo	214	473

Tabla 5.5: Matriz de Regresión Logística. Fuente: Elaboración propia.

	Predicción negativa	Predicción positiva
Verdadero negativo	478	209
Falso negativo	293	394

Finalmente se buscó la importancia de cada variable en cada modelo trabajado. En el caso del modelo SVM las preferencias fueron las siguientes:

Tabla 5.6: Importancia de las variables en el modelo SVM. Fuente: Elaboración propia.

Variable	Importancia
Cargo_3	1.999062
Proporcion	0.007679
Carrera_GL	0.003289
Carrera_MI	0.002667
Carrera_IN	0.002583
Género_F	0.002175
CursosAnteriores	0.001407
Carrera_MA	0.001111
Genero_profesor_F	0.000886
Nota_con_la_que_aprobo	0.000638
Toma_Pretoma	0.000477
SemestreCargo	0.000423
Generos_iguales_Si	0.000377

En el modelo SVM, se encontró que la variable más importante es "Cargo\_3", con un coeficiente de 1.99, lo que indica que la postulación para ser auxiliar tiene una influencia significativa en la selección de estudiantes. Además, se observó que variables como "Proporcionz Carrera.GL" también tienen cierta importancia, aunque en menor medida.

Por otro lado para el modelo XGBoost son estas:



Tabla 5.7: Importancia de las variables en el modelo XGBoost. Fuente: Elaboración propia.

Variable	Importancia
Cargo_3	0.218904
Carrera_MI	0.175669
Proporcion	0.082358
Nota_con_la_que_aprobo	0.078565
Generos_iguales_Si	0.063991
SemestreCargo	0.055722
Genero_profesor_F	0.053438
Toma_Pretoma	0.047526
Carrera_GL	0.045933
CursosAnteriores	0.043471
Carrera_IN	0.042078
EdadCargo	0.038824
Género_F	0.031043
Carrera_MA	0.022478

Por otro lado, en el modelo XGBoost, se determinó que “Cargo\_3” y “Carrera\_MI” son las variables más relevantes, con coeficientes de 0.22 y 0.18, respectivamente. Estos resultados indican que el cargo ocupado y la carrera del estudiante, en este caso minería son factores cruciales en el proceso de selección, pero lo interesante es ver que no para todos los modelos la importancia de las carreras es igual.

Ahora el resultado de importancia de variables para el modelo Random Forest son las siguientes:

Tabla 5.8: Importancia de las variables en el modelo Random Forest. Fuente: Elaboración propia.

Variable	Importancia
Nota_con_la_que_aprobo	0.255566
Proporcion	0.187098
CursosAnteriores	0.175431
EdadCargo	0.124582
Cargo_3	0.075800
SemestreCargo	0.037341
Genero_profesor_F	0.026679
Generos_iguales_Si	0.024526
Toma_Pretoma	0.022550
Carrera_MI	0.021340
Carrera_IN	0.019528
Género_F	0.013114
Carrera_GL	0.011563
Carrera_MA	0.004882

Finalmente se obtuvo los coeficientes de las variables para el modelo regresión logística:

Tabla 5.9: Coeficientes de las variables en el modelo de regresión logística. Fuente: Elaboración propia.

Variable	Coficiente
Nota_con_la_que_aprobo	0.884543
CursosAnteriores	0.601157
EdadCargo	0.568521
Género_F	0.053318
Generos_iguales_Si	0.014130
Toma_Pretoma	-0.101482
SemestreCargo	-0.268006
Carrera_IN	-0.286357
Carrera_MA	-0.311532
Genero_profesor_F	-0.343700
Carrera_GL	-0.441512
Carrera_MI	-0.696431
Proporcion	-0.726755
Cargo_3	-0.959557

En el modelo de regresión logística, se encontró que varias variables tienen una influencia significativa en el proceso de selección de estudiantes. La variable más importante es "Nota\_con\_la\_que\_aprobo", con un coeficiente de 0.884543. Esto sugiere que la calificación obtenida por el estudiante es un factor crucial en su selección.

Además, variables como "CursosAnteriores", "EdadCargo" y "Género\_F" también tienen coeficientes positivos, lo que indica que desempeñan un papel relevante en la selección. Estos resultados sugieren que el rendimiento académico previo, la experiencia previa en cursos y la edad del estudiante en relación con su cargo influyen en su probabilidad de ser seleccionado, pero en el caso del género este coeficiente si bien es positivo, es mucho más bajo que los antes nombrados.

# Capítulo 6

## Discusión de resultados

En primer lugar de la literatura se logra desprender la importancia de los modelos de rol en las carreras STEM para luchar con las brechas de género, lo que en gran medida justifica la importancia de estudiar las participaciones de las mujeres no sólo en los cargos más importantes de la universidad, como las docentes, sino que en distintos espacios de exposición, como en este caso los equipos docentes.

Del análisis exploratorio y descriptivo de datos se logró observar varias características importantes de los postulantes y seleccionados. En primer lugar, de los gráficos 4.1.1 se puede observar que en las carreras ingeniería eléctrica y minería desde el año 2020, para ambos géneros, la cantidad de estudiantes aceptados es menor, lo que podría significar que los cargos para equipos docentes en esas carreras están disminuyendo.

Algo interesante de observar también es que en las carreras Minería eléctrica e industrias se observó un peak de crecimiento en los cargos docentes seleccionados en los años 2018 y 2019, pero llegando al 2020 esto cae, lo cual coincide con el periodo de pandemia, lo cual lleva a cuestionarse cuál es el efecto detrás de esto ¿Las mujeres tuvieron una motivación a postular mayor en el año 2018 para luego presentar un periodo de resguardo ante estos cargo en periodo de pandemia o es estos fenómenos tuvieron algún efecto en la selección de los profesores?

Esto se puede comprobar con los gráficos 4.1.2 en donde se puede observar la evolución de los postulantes y aceptados por género en cada carrera.

Como se mencionó anteriormente en caso de existir un sesgo de selección por género se esperaría un comportamiento contrario entre curvas, lo cual no es el caso, lo que daría un buen primer indicio de que no existe tal sesgo. Lo interesante de estas gráficas es observar que efectivamente las mujeres bajaron su participación en el proceso de postulación y la curva decae de una forma mucho más drástica que la recta de las estudiantes aceptadas, lo que significa que el fenómeno se debió a una desmotivación de postulantes.

Otro fenómeno interesante es notar que la curva para las mujeres de la carrera ingeniería en matemáticas nunca decae en ninguno de los casos, por lo que, si bien son 3 estudiantes apenas las que participaron de estos cargos en 2021, se podría decir que existe la disposición

de trabajar con ellas de parte de los profesores, tan así que de hecho en ese año todas quienes postularon fueron aceptadas.

En el caso de los hombres se puede observar en el caso de ingeniería en minas que ha habido un decrecimiento sostenido en el tiempo hasta el año 2020 a diferencia del comportamiento de las mujeres que ha sido más cambiante, pero en el caso de ellos la selección no ha sido muy alterada, es casi una línea horizontal, lo que lleva a cuestionarse qué está ocurriendo con la motivación de estos estudiantes a participar en tales cargos. Como se mencionó anteriormente en la presentación de resultados en general las curvas se mueven de formas muy similares, pero a diferencia de las mujeres las únicas carreras que en periodo de pandemia presentaron también una caída en las postulaciones fueron industrias y eléctrica, en los otros casos esto no varió e incluso en matemáticas hubo un aumento notable.

Otro fenómeno interesante de observar es lo que ocurre con las experiencias de los cargos docentes que presentaban los estudiantes a la hora de postular, como se observa en las gráficas del ítem 4.1.4. para las carreras geología e industrias el promedio de ellas es mayor para las mujeres tanto postulantes como aceptadas, lo que permite inferir que en general las personas que participan o desean hacerlo son quienes ya han vivido una experiencia como equipo docente, lo que podría ser un gran llamado de atención para lo que es el proceso de llamado a participar, pues al parecer este no estaría llegando por completo a todas las mujeres y esto se ve incluso aún en los casos en los que la experiencia de los hombres es mayor, como en las carreras minería, y matemáticas, dado que las diferencias en general entre ellos no es muy alta.

Algo que es muy llamativo también de estos gráficos es el caso de lo ocurrido en matemáticas en donde las mujeres aceptadas presentaban menos experiencia que las rechazadas, lo cual llevaría a creer que existe una decisión de los profesores (no se sabe si de forma voluntaria o involuntaria) de hacer un recambio en las personas que participan e incluir más a mujeres que postulan pero no habían participado tan intensamente de ciertos cargos, lo mismo ocurre para los hombres pero de una forma mucho menos notoria, lo mismo ocurre para los hombres de la carrera eléctrica, no así para sus estudiantes mujeres.

Por otro lado, es interesante notar en el caso de los gráficos 4.5.1 en donde se observa si el estudiante tuvo un profesor auxiliar de su mismo sexo a la hora de postular a un cargo, y esto es la pequeña diferencia que existe en la carrera de geología, esto significa que el modelo de rol en esta carrera puede que tenga un efecto mayor que en las otras.

Ahora al observar los gráficos de densidad estudiados en el ítem 4.1.6 se pudo abstraer una información importante para las 6 carreras y es que las mujeres aceptadas en su mayoría tuvieron un mejor rendimiento académico a la hora de cursar el ramo del cual postularon, lo que podría deberse a dos cosas, una que los profesores se fijen más a la hora de aceptar a un estudiante, en este caso mujer, en su rendimiento académico o que las mujeres con peores notas que las que postularon prefieren no postular, lo que sería un sesgo de autoselección.

Al realizar los distintos modelos y observar el resultado de sus entrenamientos, se obtiene que en general los modelos tienen un rendimiento con un recall ponderado desde el 60 %, pero en el caso de los modelos XGBoos y random forest esto supera el 70 %. Estos porcentajes representan la proporción de estudiantes seleccionados que fueron correctamente identificados

por los modelos como tales, por lo que a la hora de concluir es importante considerar que estos modelos no explican el 100% de los seleccionados y que podrían existir otras variables que expliquen la selección, que podría a la vez impactar en las prioridades encontradas de las variables por cada modelo. Al observar las matrices de confusión se confirma que ningún modelo es muy preciso, por ejemplo en el caso XGB que tiene el mejor recall ponderado, su matriz indica que el modelo tiene una alta tasa de verdaderos positivos y verdaderos negativos, lo cual es positivo. Sin embargo, la cantidad de falsos positivos y falsos negativos aún es relativamente alta, lo que significa que el modelo aún puede mejorar en términos de precisión y recall.

Los resultados de la importancia de las variables de cada modelo pueden ser útiles para entender cómo cada modelo está tomando decisiones y qué características son más importantes en la predicción.

En este caso, parece que hay ciertas variables que son importantes en varios modelos, como Cargo 3z "Proporción", mientras que otras variables son importantes solo en un modelo específico, como "Nota con la que aprobo." en Random Forest y Carrera MA. en Regresión Logística.

En los resultados presentados, el género parece no ser una variable muy importante en la toma de decisión de seleccionar o no a un estudiante. En la mayoría de los modelos, el género aparece como una de las variables menos importantes, con una importancia relativamente baja en comparación con otras variables

# Capítulo 7

## Conclusiones

### 7.1. Conclusiones generales del trabajo

En conclusión, el análisis exploratorio y descriptivo de los datos sugiere que el género de los postulantes no es una variable importante en la selección de los estudiantes para cargos docentes en las carreras STEM. Sin embargo, se observaron ciertos patrones interesantes en los datos que podrían ser indicativos de ciertos sesgos o fenómenos relacionados con el género.

En particular, se observó una disminución en la participación de las mujeres en el proceso de postulación durante el periodo de pandemia en ciertas carreras, lo que podría ser un indicio de desmotivación por una causante externa. También se observó que las mujeres aceptadas en su mayoría tuvieron un mejor rendimiento académico a la hora de cursar el ramo del cual postularon, lo que podría ser un indicio de que los profesores se fijan más en el rendimiento académico a la hora de seleccionar a los estudiantes o que las mujeres postulan solo cuando tienen un muy buen rendimiento en el área, lo que correspondería a un posible sesgo de autoselección.

Por otro lado, se observó que la experiencia como equipo docente previa es una variable importante en la selección de los estudiantes, lo que podría indicar que hay un sesgo de autoselección en las personas que postulan a estos cargos. Se observó que aquellos estudiantes que tenían experiencia previa como equipo docente tenían una mayor probabilidad de ser seleccionados para los cargos que aquellos que no tenían experiencia previa. Esto podría indicar que las personas que ya han tenido experiencia previa en el cargo tienen más confianza y habilidades para realizar el trabajo, lo que les da una ventaja en el proceso de selección. Sin embargo, también es posible que haya un sesgo en el proceso de selección, donde los profesores prefieren seleccionar a estudiantes que ya tienen experiencia previa, lo que podría excluir a aquellos que no han tenido la oportunidad de obtener experiencia.

En cualquier caso, este sesgo de autoselección podría estar limitando la diversidad y la equidad en la selección de estudiantes para estos cargos. En cuanto a los modelos predictivos, se encontró que el rendimiento de los modelos fue relativamente bueno, con un recall ponderado del 60-70%. Sin embargo, se observó que ningún modelo fue muy preciso y aún hay margen de mejora en términos de precisión y recall.

Es importante tener en cuenta que el análisis realizado en este estudio es limitado y puede haber variables no consideradas que podrían tener un impacto en la selección de los estudiantes.

Es necesario seguir investigando y analizando los datos para comprender mejor los fenómenos y factores que influyen en la selección de estudiantes para el equipo docente, y tomar medidas para garantizar una selección objetiva e imparcial en términos de género y otros factores. También es importante considerar medidas para fomentar la participación de mujeres en los equipos docentes de las carreras STEM para evitar el círculo vicioso que puede surgir cuando hay menos mujeres en estos roles. Es importante considerar cómo estos modelos de rol influyen en la percepción que los estudiantes tienen sobre sus habilidades y posibilidades en una carrera STEM. Si se identifica una falta de perspectiva de género en la elección de carreras STEM, sería necesario tomar medidas para fomentar una perspectiva más amplia y equitativa, que permita a los estudiantes elegir el camino en su carrera basándose en sus intereses y habilidades, independientemente de su género.

## **7.2. Recomendaciones para la facultad y trabajos futuros**

Es importante continuar con las políticas que se han generado con el fin de promover el ingreso de mujeres en la facultad, lo que como se puede ver en las gráficas sin duda ha funcionado, pero también es necesario seguir haciendo un llamado general a la participación de las mujeres en distintos espacios importantes que pueden generar un desarrollo educativo y personal para los estudiantes, como en este caso los equipos docentes.

Sería recomendable que los profesores y cargos docentes durante el semestre promuevan los beneficios de participar en estas iniciativas u otras similares. Además, es importante darles la oportunidad a distintos talentos, por lo que hay que intentar no promover tanta recurrencia de los estudiantes como docente del mismo ramo, esto con el fin de también evitar el sesgo de autoselección antes nombrado.

Se recomienda además seguir investigando la real participación de las mujeres en estos espacios educativos y sociales como lo son los trabajos dirigidos, centros de estudiantes, intercambios y prácticas.

Además, para analizar de mejor forma un sesgo de autoselección se recomienda replicar el estudio, pero esta vez para predecir si la disposición de los estudiantes a postular. Por otro lado, para analizar de mejor forma los modelos de rol sería interesante estudiar las trayectorias de las estudiantes en la fcfm y observar todos los modelos en cargos importantes que han tenido, no solo profesores, sino que también cargos docentes, puesto que ellos incluso son más cercanos en muchos casos que los profesores.

Si se pudiese repetir este estudio o para alguno futuro, se recomendaría trabajar con la base completa de la universidad y no seleccionar solo algunas carreras, además se recomendaría ampliar el rango de tiempo estudiado, puesto que se pudo observar que cuándo ocurren sucesos importantes como lo son la pandemia, el estallido social y la toma feminista, existe

un cambio en el comportamiento de las postulaciones.

Finalmente se recomienda complementar este estudio con análisis cualitativos tales como entrevistas a estudiantes, auxiliares, ayudantes y profesores, con el fin de confirmar lo observado por los datos.



# Bibliografía

- [1] Desarrollo docente old eic. <https://ingenieria.uchile.cl/escuela/docencia/desarrollo-docente>.
- [2] Hechos y cifras eic. <https://ingenieria.uchile.cl/escuela/la-escuela/hechos-y-cifras>.
- [3] Hechos y cifras fcfm. <https://ingenieria.uchile.cl/sobre-la-fcfm/hechos-y-cifras.html>.
- [4] Nuestra escuela. <https://ingenieria.uchile.cl/escuela/la-escuela/presentacion.html>.
- [5] Organigrama eic. <https://ingenieria.uchile.cl/escuela/la-escuela/estructura/organigrama.html>.
- [6] Organigrama fcfm. <https://ingenieria.uchile.cl/sobre-la-fcfm/organizacion/personas/organigrama.html>.
- [7] Sobre la fcfm. <https://ingenieria.uchile.cl/sobre-la-fcfm>.
- [8] *Mujeres en STEM: cómo romper con el círculo vicioso*. CIPPEC, 2020.
- [9] Joshua D Angrist and Jorn-Steffen Pischke. *Mostly Harmless Econometrics*. Princeton University Press, 2008.
- [10] N. Bonilla. Memoria para optar al título de ingeniero civil industrial: Experiencia académica de estudiantes mujeres en primer año de ingeniería. propuestas de mejoramiento de los sistemas de apoyo a estudiantes con mirada de género. <https://repositorio.uchile.cl/bitstream/handle/2250/142486/Experiencia-academica-de-estudiantesmujeres-en-primer-ao-de-ingenieria-Propuestaspdf?sequence=1&isAllowed=y>, 2016.
- [11] A. Canales, M.I. Cortez, M. Sáez, and A. Vera. *Brechas de género en carreras STEM*. Concurso Políticas Públicas UC, 2021.
- [12] F. Chollet. *Deep learning with Python*. Manning Publications Co., 2018.
- [13] P. Cifuentes. *Brecha de género en Ciencias, Tecnología, Ingeniería y Matemáticas (STEM): Aspectos teóricos y experiencia extranjera*. Biblioteca Del Congreso Nacional de Chile, 2020.

- [14] Florencia Antonia Correa Campos. Análisis de diferencias de género en la evaluación docente de los cursos dictados por el departamento de ingeniería civil industrial. Master's thesis, Universidad de Chile, Santiago de Chile, 2019.
- [15] F. A. Cuevas. Participación de mujeres en stem, la situación chilena y comparación internacional. Recuperado de <https://accioneducar.cl/participacion-de-mujeres-en-stem-la-situacion-chilena-y-comparacion-internacional> 2022.
- [16] DDG FCFM. Observatorio fcfm. <https://ddg.ingenieria.uchile.cl/observatorio-fcfm/>, s.f.
- [17] S. De La Fuente. *Análisis de Conglomerados*. UAM, Madrid, 2011.
- [18] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. The kdd process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11):27–34, 1996.
- [19] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. The kdd process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11):27–34, 1996.
- [20] FCFM. Normas generales sobre la participación de auxiliares y ayudantes en la docencia de pregrado de la fcfm, 2016.
- [21] C. González-Ugalde. Investigación fenomenográfica. *Revista Internacional de Investigación en Educación*, 7(14):141–158, 2014.
- [22] J. González. *Algoritmos de Aprendizaje Supervisado en la Clasificación de Exoplanetas en Python*. Universidad Antonio Nariño, 2021.
- [23] J. Hermanowicz. The great interview: 25 strategies for studying people in bed. *Qualitative Sociology*, 25(4):479–499, 2002.
- [24] R. Hernández. *Metodología de la investigación*. McGraw Hill Education, 6ta edición edition, 2014.
- [25] Instituto Nacional de Estadísticas. *MUJERES EN CHILE Y MERCADO DEL TRABAJO. Participación laboral femenina y brechas salariales*. Instituto Nacional de Estadísticas, 2015.
- [26] C. Jorquera. *Modelo STEM y equidad de género*. Biblioteca Del Congreso Nacional de Chile, 2020.
- [27] A. Kassambara. *Practical Guide to Cluster Analysis in R*. STHDA, 2017.
- [28] J. Kim and S. Celis. Women in stem in chilean higher education: Social movements and institutional transformations. In H. K. Ro, F. Fernandez, and E. J. Ramon, editors, *Gender Equity in STEM in Higher Education: International Perspectives on Policy, Institutional Culture, and Individual Choice*, pages 105–120. 2021.

- [29] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence*, 14(2):1137–1145, 1995.
- [30] M. Lamas. La perspectiva de género, s.f.
- [31] S. Penroz. Memoria para optar al título de ingeniera civil industrial: Análisis de la continuidad de estudios entre pre y post grado de estudiantes de la facultad de ciencias físicas y matemáticas de la universidad de chile, con enfoque de género. Obtenido de <https://repositorio.uchile.cl/bitstream/handle/2250/178753/Analisis-de-la-continuidad-de-estudios-entre-pre-y-post-grado-de-estudiantes-de-1.pdf?sequence=1&isAllowed=y>, 2020.
- [32] E. Piña and C. Fernandez. *Propuestas para reducir las brechas laborales de género en Chile*. Centro Políticas Públicas UC, 2021.
- [33] John. A. RICE. *Mathematical Statictcs and Data Analysis*. Editorial Thomson Brooks/Cole, 2006.
- [34] John A. Rice. *Mathematical Statistics and Data Analysis*. Thomson Brooks/Cole, 2006.
- [35] J. Tukey. *Exploratory Data Analysis*. Addison-Wesley Series in Behavioral Science: Quantitative Methods. 1997.
- [36] John Tukey. *Exploratory Data Analysis*. Addison-Wesley Series in Behavioral Science: Quantitative Methods, 1997.
- [37] UNESCO. *Informe de Seguimiento de la Educación en el Mundo, 2016: La Educación al servicio de los pueblos y el planeta: creación de futuros sostenibles para todos*. UNESCO, 2017.
- [38] Universidad de Chile. Estatuto de la universidad de chile: Título iv. de los académicos. Obtenido de <https://uchile.cl/presentacion/institucionalidad/estatuto-de-la-universidad-de-chile/titulo-iv-de-los-academicos>.
- [39] G. Valenzuela González. Aprendizaje supervisado: Métodos, propiedades y aplicaciones. Universidad de Malaga, 2022.