

PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR
FACULTAD DE ECONOMÍA

Disertación previa para la obtención del título de Economista

***Propuesta alternativa para la categorización
socioeconómica de los estudiantes de la PUCE-Q, año
2018***

Dayana Camila Salazar Soria
dsalazar163@puce.edu.ec

Director: Ec. Abner Bravo H., Mtr.
abravo@puce.edu.ec

Quito, junio de 2020

Resumen

El Estado ecuatoriano en su normativa educativa, busca garantizar el acceso a la educación, sin discriminación de ningún tipo y bajo el principio de igualdad de oportunidades. Por lo mismo, en el caso de la educación superior, el órgano estatal guía a las universidades y escuelas politécnicas a conservar este objetivo incentivando la implementación de sistemas de beneficios o ayudas económicas. En la misma línea, se indica que es parte de la autonomía universitaria el asignar estos beneficios según su estructura financiera, gestión y criterio. La Pontificia Universidad Católica del Ecuador (PUCE) se rige a esta normativa a través del Sistema de Beca Socioeconómica (SBS) que categoriza a los estudiantes bajo principios de solidaridad, equidad, entre otros. Este análisis se enfoca en sugerir una propuesta alternativa de clasificación con el uso de la técnica algoritmo k-medias de minería de datos. Los hallazgos principales de la aplicación de este método indican una clasificación sesgada hacia una de las categorías analizadas, además de mostrar resultados diferentes al modelo PUCE, que presenta una agrupación más proporcionada entre categorías.

Palabras clave: Clasificación Socioeconómica, Modelo de Categorización, Acceso a Educación, Becas, Algoritmo k-medias.

Abstract

The Ecuadorian government in its educational regulations seeks to ensure access to education, without discrimination of any kind and under the principle of equal opportunities. In the case of higher education, the state body guides universities and polytechnic school to maintain this objective by incentivizing the implementation of benefit systems or financial aid. In the same vein, it is indicated that it is part of university autonomy to allocate these benefits according to their financial structure, management and criteria. The Pontificia Universidad Católica del Ecuador (PUCE) adheres to these regulations through the Sistema de Beca Socioeconómica (SBS) which categorizes students under principles of solidarity, equity, among others. This analysis focuses on suggesting an alternative classification proposal with the use of the k-means data mining algorithm technique. The main findings of the application of this method indicate a biased classification towards one of the analyzed categories, in addition to showing different results to the PUCE model, which presents a more proportionate grouping between categories.

Keywords: Socioeconomic Classification, Categorization Model, Access to Education, Scholarships, K-means Algorithm.

A Dios por darme lo que necesito y cuando lo necesito, por su fuerza y bondad.

A mis padres por su amor, paciencia, sabiduría y sobre todo por su apoyo incondicional durante esta y todas las etapas de mi vida. Por demostrarme con el ejemplo que el sacrificio y trabajo dan sus frutos. Por creer en mí, impulsar mis metas, motivar mis sueños y por sacrificar tanto por mi futuro y bienestar. El esfuerzo realizado en mi época universitaria no se compara con la lucha que han realizado en sus vidas, y por lo mismo son mi más grande admiración.

A David: te quiero, te extraño y te pienso todos los días.

Un agradecimiento especial al Ec. Abner Bravo por ayudarme en el desarrollo de esta disertación, por haber guiado y centrado mis ideas con paciencia y bondad. Por su tiempo y sabiduría. Aprecio de sobremanera sus aportes, los cuales siempre me brindaron una nueva perspectiva para analizar las cosas. Por darme una oportunidad para aprender a su lado, por siempre confiar en mí, y sobre todo en esta investigación.

A mis profesores, que durante mi vida universitaria aportaron desde la academia y su experiencia un nivel de conocimiento inigualable.

A mis compañeros por apoyarnos durante este camino, por los consejos, la ayuda y los buenos momentos. En especial a Gaby por el apoyo moral y su valiosa amistad.

A mi familia por el apoyo, a mis hermanas por ser el mejor ejemplo.

A Les y Shey por estar siempre en las buenas y en las malas, por su incondicional amistad y apoyo en todos estos años.

Propuesta alternativa para la categorización socioeconómica de los estudiantes de la PUCE-Q, año 2018

Introducción.....	10
Metodología del trabajo.....	12
Preguntas generales y específicas	12
Objetivos generales y específicos.....	12
Tipo de investigación	13
Diseño de la investigación	13
Técnicas y procesamiento de información	13
Fundamentación teórica.....	16
Educación.....	16
Inversión en capital humano	19
Subsidios y transferencias no monetarias.....	21
Teoría del método: algoritmo k-medias.....	22
Tipos de clústeres	24
K-medias.....	24
Capítulo I: Sistema de categorización de la PUCE-Q.....	27
Introducción.....	27
1.1. Descripción de beneficios económicos y categorías	27
1.2. Descripción general del modelo actual de categorización.....	30
1.2.1. Metodología del modelo	32
1.3. Análisis de resultados de la distribución de estudiantes según las categorías del Sistema de Beca Socioeconómica 2018.....	40
1.3.1. Análisis de resultados 2018.....	40
Capítulo II: Modelo alternativo de categorización	42
2.1. Motivación.....	42
2.2. Descripción general del modelo alternativo.....	43
2.2.1. Procesamiento de la base de datos.....	43
2.2.2. Transformación y selección de variables	44
2.3. Aplicación del análisis Clúster con data 2018.	47
2.3.1. Aplicación del algoritmo k-medias con variables cuantitativas	53
Capítulo III: Incidencia de los resultados del modelo alternativo.....	56
3.1. Contraste entre los resultados de categorización del modelo alternativo con resultados del Sistema de Beca Socioeconómica 2018.....	56

3.2. Consideración de otros factores que afectan a la asignación de recursos con el modelo de categorización	60
Conclusiones.....	62
Recomendaciones.....	63
Referencias Bibliográficas	64
Anexos	68

Índice de cuadros

Cuadro 1: Asignación de categoría según rango de puntuación del modelo.....	30
Cuadro 2: Componentes de cada dimensión del modelo de categorización de la PUCE-Q..	33
Cuadro 3: Ponderación de las dimensiones y subdimensiones del modelo de categorización de la PUCE-Q.	39
Cuadro 4: Clasificación de variables del modelo según uso en los modelos.	43
Cuadro 5: Operacionalización de variables.	45
Cuadro 6: Variables del modelo, tipo y denominación en R.....	47
Cuadro 7: Número de estudiantes por categoría según modelo cuantitativo/cualitativo.....	49
Cuadro 8: Medias por variable de cada clúster modelo cuantitativo.	50
Cuadro 9: Número de estudiantes por categoría según modelo cuantitativo.	53
Cuadro 10: Medias por variable de cada clúster modelo cuantitativo.	54
Cuadro 11: Comparación de estadísticos y características de los modelos alternativos.....	56
Cuadro 12: Resultados por número de estudiantes en cada categoría.....	56
Cuadro 13: Resultados por proporción de agrupación de categorías.	59
Cuadro 14: Costo económico aproximado de los modelos.	61

Índice de gráficos

Gráfico 1: Porcentaje de estudiantes de la PUCE-Q por tipo de beca total o parcial (2018). 28	28
Gráfico 2: Visualización de categorías según el descuento recibido (2018).....	29
Gráfico 3: Proceso de uso de información para asignación de categoría SBS.	32
Gráfico 4: Ingreso familiar de los estudiantes de la PUCE-Q (2018).....	35
Gráfico 5: Cantidad de miembros en el hogar de los estudiantes de la PUCE-Q (2018).....	36
Gráfico 6: Porcentaje de estudiantes con discapacidad de la PUCE-Q por tipo de discapacidad (2018).	36
Gráfico 7: Estudiantes de la PUCE-Q con enfermedad catastrófica (2018).	38
Gráfico 8: Estudiantes de la PUCE-Q por tipo de colegio (2018).....	38
Gráfico 9: Distribución de estudiantes por categoría – PUCE-Q (2018).....	40
Gráfico 10: Distribución de estudiantes por categoría- Facultad Economía (2018).....	41
Gráfico 11: Distribución de ingresos por clúster del modelo cuantitativo/ cualitativo.	51
Gráfico 12: Distribución de gastos por clúster del modelo cuantitativo/ cualitativo.....	51
Gráfico 13: Porcentaje de gasto total sobre el ingreso total del modelo cuantitativo/ cualitativo.	52
Gráfico 14: Clasificación de estudiantes en las cinco categorías por modelo.	57
Gráfico 15: Clasificación de estudiantes por categoría: modelo PUCE.....	57
Gráfico 16: Clasificación de estudiantes por categoría: modelo Cualitativo/ Cuantitativo.....	58

Gráfico 17: Clasificación de estudiantes por categoría: modelo Cuantitativo.	58
Gráfico 18: Diferencias en categorías asignadas: modelo PUCE vs modelos alternativos. ..	59

Introducción

La Pontificia Universidad Católica del Ecuador (PUCE), como institución de educación superior cofinanciada, recibe tanto recursos privados como públicos. Esto debe de considerarse al momento de asignar los mismos a diversas actividades de gestión, académicas, financieras, etc. Por lo mismo, y como se recoge en la Ley Orgánica de Educación Superior (LOES), las instituciones deben redestinar la totalidad de estos montos a ayudas económicas conocidas como becas o créditos educativos.

Parte de esta lógica, y según el artículo 2 de la LOES, involucra que se debe hacer esta distribución bajo los principios de solidaridad, igualdad de oportunidades, equidad, entre otros. Y es que no solo la PUCE se rige a esto, sino también otras universidades de la región. Específicamente, miembros de la Asociación de Universidades Confiadas a la Compañía de Jesús en América Latina (AUSJAL) de la cual la PUCE forma parte (LOES, 2018).

El proceso que resume esta asignación es el Sistema de Beca Socioeconómica (SBS). El sistema referenciado cuenta con un modelo que permite clasificar estudiantes según sus condiciones sociales y económicas para, en base a este criterio asignar una beca o ayuda económica. En relación a la problemática de asignación de recursos, es pertinente que este modelo evolucione con el fin de mejorar sus resultados de categorización.

La administración de la PUCE usa un modelo de categorización en el que asigna cierto porcentaje de descuento en el pago total de acuerdo a la categoría asignada. Para el año 2018, de un total de 10.496 estudiantes de la sede Quito constan, 1.954 estudiantes en la categoría A, 2.845 en la categoría B, 2.506 en la categoría C, 1.119 en la categoría D y 2.072 en la categoría E. Esta última no recibe ningún beneficio económico en el pago de su matrícula y créditos, mientras que A recibe el mayor monto de descuento en su pago (Dirección General de Estudiantes, 2018).

La importancia de la investigación recae en aplicar una metodología alternativa a la aplicada por la PUCE para la asignación de la beca socioeconómica. Pues, de una categoría asignada de manera correcta (acorde a factores sociales y económicos del estudiante) depende, en gran parte, la continuidad en estudios, la calidad y desarrollo académico de los estudiantes que se convertirán en profesionales. Asimismo, esto influenciaría a una mejor distribución de los recursos de los estudiantes, minimizando las desigualdades que se pueden presentar cuando los mismos se asignan. Básicamente, como indica el óptimo de Pareto, mejorando el bienestar de una persona sin perjudicar el de otra. (Pérez, Laguado, & Martínez)

Con el objetivo de probar una nueva metodología de clasificación, el primer capítulo desarrolla a detalle los componentes del modelo actual del SBS y evidencia los resultados del mismo. Lo anterior con fuente de información de la Dirección de Beneficios Económicos (DBE) y sus reportes del proyecto de Beca Socioeconómica.

Posteriormente, en el segundo capítulo, se procede a desarrollar el modelo alternativo de clasificación con una técnica diferente. El algoritmo k-medias se expone con la concepción teórica de los autores del área de minería de datos, Daniel Peña y Pang-Ning Tan. Este algoritmo es una herramienta que en la mayoría de los casos trabaja con variables cuantitativas, y dada la importancia de también analizar variables cualitativas que recogen características sociales del entorno del estudiante, se obtiene un modelo cuantitativo/cualitativo y un modelo cuantitativo.

Cabe destacar que la clasificación resultante alcanza un óptimo en la metodología, no necesariamente un óptimo económico relacionado con eficiencia. El algoritmo k-medias, comprende una asignación de la observación al mejor clúster mediante la maximización de su distancia euclídea. Se comprende entonces que cada estudiante estará categorizado en su óptimo.

En el último capítulo, y como una de las motivaciones principales, se procede a confrontar los resultados de clasificación de los dos modelos desarrollados junto al del SBS. Se establecen comparaciones sobre la distribución, variables y resultados generales. Además, se visualizan las evidentes diferencias de los modelos al trabajar con metodologías distintas. Cabe aclarar que no se busca evaluar el modelo actual de categorización de la PUCE, sino el comparar el resultado del mismo con lo que arroje un modelo alternativo realizado con otra técnica.

Se incluye un análisis sobre el costo económico de cada modelo para entender, desde otro punto de vista, como el modelamiento puede generar un ahorro de recursos debido a un correcto direccionamiento de su presupuesto. Finalmente, se recogen las reflexiones del trabajo de disertación según sus puntos más importantes.

Metodología del trabajo

Preguntas generales y específicas

- ***General***

¿Qué propuesta alternativa de categorización socioeconómica puede sugerirse para los estudiantes de la PUCE, para el año 2018?

- ***Específicas***

¿En qué consiste el Sistema de Beca Socioeconómica de la PUCE-Q?

¿Qué modelo de categorización socioeconómica alternativo puede sugerirse para los estudiantes de la PUCE?

¿En qué difieren los resultados de categorización socioeconómica de los estudiantes, obtenidos con un modelo alternativo, con los del Sistema de Beca Socioeconómica?

Objetivos generales y específicos

- ***General***

Proponer una alternativa de categorización socioeconómica para los estudiantes de la PUCE, para el año 2018.

- ***Específicos***

Describir el Sistema de Beca Socioeconómica de la PUCE-Q.

Sugerir un modelo de categorización socioeconómica alternativo para los estudiantes de la PUCE.

Contrastar los resultados de categorización socioeconómica de los estudiantes, obtenidos con un modelo alternativo, con los resultados del Sistema de Beca Socioeconómica.

Tipo de investigación

La presente disertación tiene un alcance descriptivo pues se busca especificar rasgos importantes en un grupo, en este caso, los estudiantes de la PUCE-Q. Por lo mismo, la descripción de características de esta muestra, sirven de pauta para el análisis de categorización de la beca socioeconómica que será la base para la comparación con los modelos alternativos resultantes del uso de una nueva técnica. Esto en concordancia con los objetivos de la investigación (Hernández, Fernández, & Baptista, 2010).

Diseño de la investigación

Para la definición del diseño de investigación es pertinente indicar la unidad de análisis sobre la cual se trabajó. Para esto se incurre en la definición de un universo; en este caso la totalidad de estudiantes de la Pontificia Universidad Católica del Ecuador Sede Quito (PUCE-Q); la muestra, los estudiantes de la PUCE-Q de todos los niveles y carreras que reciben beca socioeconómica; y finalmente la unidad; los estudiantes de PUCE-Q que ingresaron en los semestres de 2018 y que reciben beca socioeconómica.

Esta unidad de análisis es vista como una instancia de un fenómeno más grande, como pueden ser los estudiantes de las sedes de la PUCE o inclusive los que reciben otro tipo de beca. El escoger esta unidad corresponde a la facilidad del análisis en cuestión de datos, es decir, corresponde a la amplitud de la evidencia. El realizar este estudio en los estudiantes de nuevo ingreso de la PUCE-Q, es menos pesado que realizarlo a nivel general, con estudiantes de niveles superiores que ingresaron con otra metodología del Sistema de Beca Socioeconómica (Gerrings, 2004).

En la misma línea, esto permite que el alcance de la propuesta de investigación tenga una mayor profundidad. En perspectiva, la reducida unidad de análisis establecida recaba un estudio más detallado y centrado, en comparación a realizarlo con una muestra más grande. Como lo indica Gerrings, permite saber más de menos cosas. En este caso, conocer de manera integral el funcionamiento del sistema de asignación de la beca socioeconómica a los estudiantes de la PUCE-Q.

Por lo tanto, se trata de un Caso de Estudio II pues corresponde al análisis dentro de una unidad, con los casos mencionados anteriormente, y es de manera sincrónica pues se desarrolla una sola vez en el tiempo, año 2018, esto debido a la disponibilidad de datos (Gerrings, 2004).

Técnicas y procesamiento de información

El uso de la técnica k-means se justifica por el objetivo que persigue como método de clasificación, que permitirá encontrar semejanzas en la muestra de estudiantes de la PUCE y generará una categorización alternativa. El análisis de conglomerados permite “buscar grupos latentes en un conjunto de datos, dadas características importantes” (Attewell, Monaghan, & Kwong, 2015) esto indica que elementos similares de los datos se asociarán en diversas clasificaciones.

El trabajo de investigación de Luis Carrasco para la obtención de su título de maestría se trata de utilizar el Análisis de Componentes Principales (ACP) un Análisis Factorial de Correspondencias Múltiples que incide en la clasificación socioeconómica de los estudiantes. En donde, utilizando el ACP clasifica las variables en seis dimensiones: población, educación, vivienda, empleo, salud e ingresos. Concluye que “los resultados del modelo muestran un nivel bajo de pérdida de información y ponderadores óptimos consistentes por la distribución similar que se observa tanto entre las categorías de las variables” (Carrasco Delhy, 2014).

En esta perspectiva, Wilson Arroyo también desarrolla una metodología de clasificación socioeconómica para los estudiantes de la Escuela Politécnica Nacional. Este análisis compara cinco técnicas: Análisis de Componentes Principales Categórico, Análisis de Conglomerados, Árbol de Clasificación, Regresión Logística Multinomial y Máquinas de Vectores Soporte. Concluye que el Árbol de Clasificación “tiene un nivel de asertividad del 97%” lo que permite una clasificación certera de componentes económicos y sociales (Arroyo Ruales, 2017).

El uso del algoritmo k-means no se limita al ámbito económico o al educativo como se indica en los trabajos de Carrasco y Arroyo. Una aplicación por parte de un banco en la provincia de Zhejiang para la clasificación de empresas del año 2006, usa la técnica para agrupar según reportes financieros, esto para determinar de manera más fácil, si se otorga o no el crédito (Qian, 2006).

En un sector diferente, como es el de la construcción, los autores buscan analizar el comportamiento de las empresas de la rama agrupándolas según indicadores financieros de liquidez, rentabilidad, solvencia y gestión. Habla sobre como el utilizar el algoritmo de k-medias permite homogeneizar las características de dichas empresas para dar paso a un análisis de competencia con diferentes índices de concentración económica (Horna, Guachamín, & Guerrero, 2009).

El desarrollo de la propuesta del modelo alternativo con la técnica k-medias comenzó con el trabajo de la base de datos requerida a Dirección de Beneficios Económicos (DBE). Después de tratarla se definieron 20 variables que engloban a las indicadas en el plan de disertación. Se obtuvieron dos modelos alternativos, uno solo cuantitativo y otro cuantitativo/cualitativo.

En segundo lugar se realizan los fundamentos teóricos de la disertación con una inclinación hacia revisar la normativa nacional de educación, la Ley Orgánica de Educación Superior (LOES). Asimismo, se investiga el trato y fundamentos de otras universidades en la región con respecto a su sistema de becas utilizando documentos publicados por dichas instituciones de educación superior.

Posteriormente, es estudiado el modelo de categorización del Sistema de Beca Socioeconómica de la PUCE-Q. Se trabaja con documentos de la DBE, así como con recursos de otras sedes y de la página web de la PUCE-Q. Se resaltan las características principales del mismo.

Con la información de todos los modelos se procede a comparar los resultados de categorización. Principalmente, se usan gráficas para demostrar las diferencias en los mismos. De igual manera, se compara el número de estudiantes que está en cada categoría según el modelo y la técnica aplicada. Se incluye un análisis sobre el costo económico aproximado de cada modelo.

Con lo mismo, se generan conclusiones sobre las características de los modelos, sus principales resultados y el aporte del algoritmo k-medias. Además se incluye, el cómo los

modelos alternativos toman lo mejor del modelo de categorización de la PUCE-Q y de las ventajas del algoritmo k-medias.

En las recomendaciones se incluyen los factores que inciden en la asignación de recursos y que podrían incluirse para mejorar el modelo. Se habla también de otras variables que podría ser consideradas y de sobre como un mejor manejo de registros puede aportar a mejores resultados en los modelos analizados.

Fundamentación teórica

Educación

Para un análisis del ámbito educativo se considera varios aspectos sociales, económicos, territoriales, de conocimiento; etc. De entre ellos, el social hace referencia a la educación como derecho humano en donde es visto como un medio para el desarrollo de capacidades del individuo. Por otra parte, en relación a lo económico, es un recurso para conseguir ingresos que mejoran la calidad de vida de la persona.

El acceso a la educación es una de las principales preocupaciones tanto en el sector privado como en el público. Con respecto al primero, las instituciones educativas autofinanciadas de todos los niveles, procuran ofrecer una educación que equipare la calidad con el precio pagado para de esta manera atraer nuevos estudiantes. Es decir, si bien su objetivo es aumentar el acceso, no lo hacen desde una óptica social, sino desde un punto de vista financiero y de rentabilidad (Tierney & Hentschke, 2007).

Por otra parte, el enfoque de acceso a educación pública viene desde una perspectiva social. Principalmente, el Estado toma como responsabilidad el asegurar educación, y aun cuando no puede llegar a la mayoría, sigue impulsando proyectos, aprobando legislaciones que permitan aumentar los niveles de educación en todos los individuos (Villanueva, 2018).

El sistema educativo ecuatoriano impulsa el acceso a la educación sin discriminación de ningún tipo, como lo indica el artículo 2 de la Ley Orgánica de Educación Superior (LOES). Como parte del mismo, el sistema de educación superior establece ciertos lineamientos para el cumplimiento de esta condición de acceso. Las universidades, institutos y conservatorios que son parte de este sistema cuentan con normativas internas y se rigen a la normativa nacional que guía su funcionamiento hacia una educación justa para todos (LOES, 2018).

La normativa nacional, LOES en su artículo 5, aboga por “acceder a una educación superior de calidad y pertinente, que permita iniciar una carrera académica y/o profesional en igualdad de oportunidades” (LOES, 2018) indicando la importancia de generar condiciones de acceso que facilite el comienzo y desarrollo de los estudios universitarios hasta su culminación.

Según el artículo 12 de la LOES, el sistema de educación superior, en concordancia con lo establecido en el “Sistema Nacional de Inclusión y Equidad Social, se rige por los principios de universalidad, igualdad, equidad, progresividad, interculturalidad, solidaridad y no discriminación” (LOES, 2018) como pauta para encaminar las acciones, actividades y decisiones de los integrantes del sistema. Estos principios tienen como base al estudiante, su entorno; y a como los mismos ayudan equilibrar las condiciones de cada estudiante cuando pasa a ser parte del sistema.

Así como los principios son una base para la gestión universitaria, se aplican también en lo referente al manejo de recursos de la institución. Legalmente, según el artículo 18 de la LOES, se establece que cada integrante del sistema de educación superior tiene la “libertad para elaborar, aprobar y ejecutar el presupuesto institucional” (LOES, 2018) esto siempre sujeto a que su distribución contribuya a cumplir los objetivos establecidos, según lo establecido en el artículo 24 de la ley mencionada.

El Estado, según el artículo 30 de la LOES, procura que las Instituciones de Educación Superior (IES) que reciben recursos financieros utilicen los mismos en diversas actividades tanto académicas como de gestión. Principalmente, en el “otorgamiento de becas totales o

parciales a estudiantes de escasos recursos económicos, en estudios de tercer nivel desde el inicio de la carrera” (LOES, 2018) pues esto concuerda con el cumplimiento de los principios en los que se basa el Sistema de Educación Superior.

En la misma línea del artículo mencionado anteriormente, la obligación del Estado de otorgar becas no interfiere en la elaboración o manejo presupuestal de las IES. Por el contrario, busca “seleccionar a los estudiantes y adjudicarles las becas, en razón de su autonomía responsable y el cumplimiento de la política pública emitida para el efecto” (LOES, 2018) entonces la distribución cumple con parámetros definidos por el Estado y a la par con las directrices de manejo de recursos de cada institución.

La normativa en su artículo 4 expresa claramente que “el derecho a la educación superior consiste en el ejercicio efectivo de la igualdad de oportunidades, en función de los méritos respectivos” (LOES, 2018). Igualmente, en el artículo 5 indica que se debe “obtener de acuerdo con sus méritos académicos becas, créditos y otras formas de apoyo económico que le garantice igualdad de oportunidades en el proceso de formación de educación superior” (art. 5). Es evidente recalcar que la motivación central es el garantizar la igualdad de oportunidades de cada estudiante.

El principio de igualdad de oportunidades, según el artículo 71 de la LOES, se basa en “garantizar a todos los actores del Sistema de Educación Superior las mismas posibilidades en el acceso, permanencia, movilidad y egreso del sistema, sin discriminación de género, credo, orientación sexual, etnia, cultura, preferencia política, condición socioeconómica, de movilidad o discapacidad” (LOES, 2018) recalcando que ninguna diferencia de tipo personal o referente al entorno económico del estudiante y sus allegados interferirá en su proceso de enseñanza de educación superior.

De acuerdo con la LOES, en su artículo 78, se estipulan tres tipos de estímulos económicos a los cuales los estudiantes pueden acceder para empezar, continuar y terminar su educación superior; becas, créditos educativos y ayudas económicas. Una beca se define como una “subvención total o parcial”, mientras que un crédito educativo se refiere a una asignación de “recursos económicos reembolsables” y por el contrario una ayuda económica es una “subvención de carácter excepcional no reembolsable” (LOES, 2018).

En referencia al tema de becas, la normativa de educación en su artículo 77 establece criterios para su entrega y manejo. Como, en el caso de que se “establecerán programas de becas completas, o su equivalente en ayudas económicas a por lo menos el 10% del número de estudiantes regulares” (LOES, 2018) esto con el objetivo de asegurar un nivel mínimo para la concesión de becas. Asimismo, se indica la importancia de criterios para la asignación de la beca, como por ejemplo “condición económica, situación de vulnerabilidad, proximidad territorial, excelencia y pertinencia” (art. 77), lo cual alude a las características académicas y socioeconómicas de cada estudiante.

Algo semejante ocurre con lo establecido en el artículo 73 de la LOES, el cual indica que el cobro de valores por conceptos de aranceles, matrículas y otros, al igual que el proceso de asignación de becas, se ciñe al principio de igualdad de oportunidades. Sumado a eso en el artículo 90 de la misma ley se indica que “las instituciones de educación superior particulares tratarán de establecer un sistema diferenciado de aranceles, que observará de manera principal, la realidad socioeconómica de cada estudiante” (LOES, 2018), recalcando lo descrito anteriormente, la relevancia de basar el otorgamiento de becas en un estudio de condiciones socioeconómicas.

De manera similar los sistemas de educación superior de la región proponen una normativa similar cuando se refiere al establecimiento de ayudas económicas. Se evidencia que otras

Instituciones de Educación Superior rigen su proceso de asignación de becas en principios y criterios apegados a una normativa interna y estatal. En la mayoría de universidades prevalece la importancia de basar estas decisiones en estudios socioeconómicos del estudiante.

Para apreciar tal consideración, por ejemplo, la Universidad de Costa Rica indica que, con los objetivos de “garantizar la democratización de la educación superior universitaria, favorecer la movilidad social, fomentar la solidaridad y el compromiso social” (Regueyra Edelman, 2010) maneja los recursos relacionados con las becas proporcionadas a sus estudiantes. Esto recalca la importancia de permitir el acceso a la educación superior apegado a principios como el de solidaridad, que “se establece como uno de los principios del sistema, el destinar los ingresos de matrícula para financiar los servicios que ofrecía dicho Sistema de Becas” (pág. 6).

A propósito de lo mismo, la Universidad Nacional de Colombia reconoce que realizar un estudio socioeconómico de los estudiantes “permite identificar grupos de estudiantes según el origen social, variable indispensable para el diseño de políticas y planes de desarrollo por parte de las universidades públicas y por el Estado” (Pérez, Laguado, & Martínez). En efecto, la importancia de una distribución socioeconómica permite gestionar con un sustento claro, las decisiones que conciernen al manejo de recursos de las universidades.

En relación a la Pontificia Universidad Católica de Perú y su sistema de ayuda económica se indica que se basa en varios principios y criterios, entre ellos “excelencia académica, solidaridad e igualdad de oportunidades, responsabilidad social, formación integral, responsabilidad del estudiante” (Pontificia Universidad Católica del Perú, 2008). Cabe destacar la afinidad con lo presentado en la normativa ecuatoriana, con relación al principio de igualdad de oportunidades, lo cual permite la continuidad en el proceso universitario de cada estudiante.

Esta serie de principios y criterios también son aplicados por las 30 universidades jesuitas en la región latinoamericana que conforman la Asociación de Universidades Confiadas a la Compañía de Jesús en América Latina (AUSJAL), de la cual la Pontificia Universidad Católica del Ecuador forma parte. Esta red de instituciones incentiva el apoyo y la colaboración en la gestión de las universidades que la conforman con base en la filosofía jesuita.

Por tal uno de sus miembros, la Universidad Católica de Córdoba ubicada en Argentina indica que las ayudas económicas otorgadas se apegan a una “política de desarrollo y del compromiso social que caracteriza a la institución con una mirada inclusiva creando oportunidades de acceso y permanencia en la educación universitaria” (Universidad Católica de Córdoba, 2018). Es decir, su enfoque también procura relacionarse a principios de solidaridad que permita un acceso a la educación sin discriminación de ningún tipo.

Otro integrante de la AUSJAL, la Universidad Iberoamericana de México se guía por la Ley General de Educación de su país. Se establece que la institución “procura facilitar las posibilidades de estudio a estos alumnos cuando no cuenten con suficientes recursos económicos” (Universidad Iberoamericana Ciudad de México, 2013) indicando que se basan en un estudio socioeconómico del estudiante para determinar una distribución de ayudas económicas que favorezcan a la población estudiantil más necesitada.

De la misma manera, la Pontificia Universidad Javeriana reconoce como limitante “la dificultad económica que enfrenta un número muy representativo de estudiantes, con capacidad académica comprobada” (Pontificia Universidad Javeriana, 2014). Lo cual explica que a veces las barreras económicas no permiten que el potencial del estudiante se desarrolle en el ámbito estudiantil y posteriormente laboral.

En la red AUSJAL la Universidad Centroamericana “José Simeón Cañas” de El Salvador la idea es similar. Procura garantizar el acceso y culminación en los estudios superiores de los estudiantes, “cuyo estudio socioeconómico les asigne la cuota mínima del sistema de pago por cuotas diferenciadas” (Universidad Centroamericana José Simeón Cañas, 2018) permitiendo que no abandonen la institución en medio del plan de estudio de su carrera por circunstancias económicas.

Resulta claro que la mayoría de sistemas educativos se rige por una normativa estatal que procura una inserción justa y en igualdad de oportunidades de los estudiantes en las instituciones que los conforman. Así pues, se establecen parámetros y criterios que guían de manera interna a universidades, institutos, etc.; a manejar una asignación de ayudas económicas que focalice a estudiantes que en verdad la requieran. Al final se refleja que estos esfuerzos se realizan con miras a evitar que diversos factores opaquen la capacidad académica que un estudiante puede demostrar con los apoyos e incentivos correctos.

Inversión en capital humano

La educación es un objetivo por alcanzar en todas las regiones del mundo. Sin importar la visión ética o moral, existe una idea común en la concepción de las personas referente a que, tanto grupos vulnerables como la población en general, merece niveles de educación. Es un factor que contribuye al desarrollo productivo y por ende personal del individuo (Nishizawa, 2002).

Es importante remarcar la diferencia entre crecimiento económico y desarrollo. El primero se refiere a la capacidad económica para generar producción y un ingreso de la misma, mientras que el otro está relacionado con la mejora de las condiciones básicas humanas y también con la forma en que se enfoca en mejorar las habilidades de las personas y los niveles de preparación relacionados con sus actividades (Szirmai, 2010).

En la misma línea, un factor que influye se refiere a la educación recibida a lo largo de la vida del individuo. Su aporte es importante en cuanto ayuda al desarrollo de destrezas así como de capacidades. El constante refuerzo sobre las mismas representa un aumento en futuras oportunidades laborales traducido en un nivel de ingresos que permitan una independencia económica.

Lo último es un factor clave en relación al crecimiento de la economía de un país. Si en base a la educación adquirida, es posible obtener una estabilidad económica potenciando capacidades en un entorno laboral, los ingresos aportan a la economía. Es decir, el individuo asume responsabilidades, como el pago de impuestos, relacionadas a un enfoque ciudadano que mantiene a la economía en rumbo (Szirmai, 2010).

En referencia a lo último, se indica que para el sector privado “la productividad es responsabilidad del empresariado”. Mientras los integrantes del sector público “pueden y deben realizar una serie de actividades importantes” (Villanueva, 2018). Esto se refiere a, entre varias, realizar reformas de leyes enfatizando en mejoras en educación

En la misma línea, la Ley Orgánica de Educación Superior (LOES) en su artículo 79 dirige sus esfuerzos en guiar la gestión de las Instituciones de Educación Superior. Además, indica que “el órgano rector de la política de fortalecimiento del talento humano coordinará con el sistema financiero el otorgamiento de créditos educativos en condiciones favorables para los estudiantes de todos los niveles de la educación superior” (LOES, 2018).

En general, las implicaciones y definiciones de educación se derivan en un sinnúmero de teorías que están más enfocadas en el ámbito educativo. En cambio, una concepción que se acerca al campo económico se refiere a como la educación es vista como inversión en capital humano. El enfoque es el individuo como receptor de conocimientos que le permitirán acceder a una variedad de oportunidades que potenciarán el obtener ingresos para acceder a un mejor estilo de vida.

Sin embargo, existe el debate sobre si la educación es un gasto o una inversión. Si bien, por el lado del gasto existe una destinación de recursos importante, se reconoce que muchas de las veces pesan más los rendimientos relacionados con mejoras en calidad de vida. Esto es secundado por la teoría del Capital Humano desarrollada por varios autores y consolidada por Gary Becker.

Entre los autores se encuentra Theodore Schultz, que expresa

Propongo tratar la educación como una inversión en el hombre y tratar sus consecuencias como una forma de capital. Como la educación viene a formar parte de la persona que la recibe, me referiré a ella como capital humano. (Schultz, 1985)

La teoría del Capital Humano, considera que, al momento de tomar la decisión de invertir o no en educación, percibe los beneficios que obtendrá en el futuro si sigue estudiando y posteriormente capacitándose y los costos de esta inversión. “Seguirá estudiando si el valor actualizado neto de los costos y de las ventajas es positivo” (Destinobles, 2006).

En efecto, como se puede apreciar, la teoría del Capital Humano “considera que el agente económico tiene un comportamiento racional, invierte para sí mismo y esa inversión se realiza en base a un cálculo” (Destinobles, 2006). Esto se refiere a que al igual que se estudian los rendimientos de capital financiero de una empresa o institución, también se replica esta técnica sobre el capital humano relacionado con los costos de educarse.

Otra reflexión de Becker sobre la influencia del capital humano en la economía:

Por otra parte, Becker indica que cuando el capital humano aumenta, gracias a la alta inversión en los sectores educativos, el retorno en la inversión de capital humano aumenta hasta que éste crece lo suficiente y encuentra un equilibrio. Es decir, se eleva el ingreso per cápita de la sociedad debido a la relación directa con el crecimiento económico y el stock de capital humano. (Becker et. al, 1990:2), citado en Cardona et al. (2007)

Por lo mismo, una asignación eficiente de recursos en educación es similar a una inversión en capital humano. Los retornos de la misma se ven en un ámbito más amplio como lo es el laboral, a través del incremento de productividad del individuo. Esto se relaciona con una externalidad positiva, ya que las personas cualificadas aportan al impulso de sectores económicos:

Para Becker [...] el capital humano juega un papel importante en el estudio del desarrollo, de la distribución de la renta, de la rotación del trabajo, entre otros. A diferencia de otros autores... que sólo veían la educación como una explicación posible del crecimiento económico, Schultz [...] y Becker [...] consolidan la teoría de capital humano, y ven de la inversión en las personas un mejor futuro y una de las formas de disminuir la pobreza. Cardona et al. (2007)

En la misma línea, el crecimiento económico se basa, en cierta parte, en los aportes de un trabajador cualificado. Cuyas habilidades pueden, por ejemplo, permitir la agilización y mejora de procesos productivos. “El conocimiento de educación de un trabajador de ciertos hechos y

conceptos lo hace un trabajador más valorable” (Blaug, 1993) es decir, la educación tiene un valor agregado económico implícito.

Al mismo tiempo, se crean desigualdades entre los individuos en base a su nivel de preparación. Esta clase de inconvenientes a largo plazo hacen que los estudiantes opten por optar acceder a becas económicas que les permitan la finalización de sus estudios. No obstante, los métodos para obtener esta ayuda pueden direccionarse a ocultar información privilegiada que concierne a la institución para la asignación de recursos.

Esta asimetría afecta a una asignación de recursos eficiente. Si una de las partes no tiene la información pertinente para este propósito se incurre en una sobre asignación o sub asignación. La primera se refiere a destinar más de lo necesario a un individuo y la segunda, por el contrario a asignar menos.

Por otra parte, la dimensión de acceso a la educación trata de conseguir que se eviten barreras sociales, culturales, económicas que generen desigualdad en los estudios. Estos rasgos de vulnerabilidad se rigen a una regulación que procura por el bienestar de los estudiantes. Las normativas como la LOES se refieren a aumentos en escolaridad mediante incentivos como becas con carácter inclusivo que no discriminan a estudiantes por su origen socioeconómico, etnia, genero, discapacidad o lugar de residencia. Cumpliendo así el objetivo de lograr que más personas accedan a una educación sin barreras o impedimentos en el camino.

Subsidios y transferencias no monetarias

Más allá del análisis de la inversión individual del capital humano que propone Becker, en el estudio de la educación con la teoría económica, se observa la influencia de los actores que son parte de este proceso. Si bien se describió la función del gobierno con respecto a su intervención en la educación mediante la LOES, es importante señalar que también juega un papel interventor económico en este aspecto.

Por situaciones como fallos de mercado la intervención del Estado, de cierta manera, se ocupa de regular o equilibrar escenarios desencadenados por asimetrías de información, competencia imperfecta, distribución desigual de la renta, etc. Lo cual se interpreta desde un análisis normativo que indica las bases para la mediación estatal o que también se refiere al deber ser. Por otra parte, el rol positivo se enfoca en los efectos de las acciones del Estado o al funcionamiento de la economía en sí (Stiglitz, 2000).

Se parte con que el Estado en su rol normativo busca subsanar fallas de mercado para alcanzar un sentido de eficiencia y bienestar en diversos ámbitos, en donde la educación no es la excepción. En gran parte, la influencia del Estado en la educación ejerce importancia en el aspecto presupuestario, donde las universidades públicas y cofinanciadas son actores que participan junto con el Estado para permitir acceso a educación terciaria y de todos los niveles.

Organismos internacionales como la Organización de Estados Americanos (OEA) indica que, entre otras, el apoyo de recursos como subsidios estatales aportan a la reducción de desigualdades en la educación. Indica además que el impacto va desde proporcionar becas a hasta financiar programas de tutorías. En un estudio de la organización mencionada, se indica que los efectos sobre la desigualdad de la entrega de subsidios no tiene tanto impacto como la aplicación de transferencias monetarias, otra herramienta fiscal (OEA, 2011).

Las transferencias monetarias buscan ayudar a hogares en extrema pobreza, y así como el programa de Bono de Desarrollo Humano aplicado en Ecuador, la entrega de esta asistencia está sujeto al cumplimiento de ciertos términos. También conocido como transferencias monetarias condicionadas, busca promover inversiones en capital humano a largo plazo. Por lo general, piden requisitos en áreas de educación y salud, incentivando a las familias a mantener a sus hijos en la escuela, así como asegurarse de que reciban servicios básicos de salud (LSE, 2010).

En efecto, este tipo de ayuda condicionada demuestra haber incrementado la asistencia y las tasas de matriculación escolar. Aunque la intención es concentrar los recursos para la entrega de la transferencia a grupos afectados, al mismo tiempo puede generar cambios en el comportamiento de los hogares con el fin de recibir la ayuda. La constante identificación y comprobación de las condiciones de la familia para evitar esta clase de inconvenientes, incurre en altos costos administrativos (LSE, 2010).

En cuestiones de teoría económica, las transferencias monetarias disminuyen marginalmente la oferta laboral. Dado que este tipo de ayuda puede crear el efecto contrario al esperado y desincentivar el trabajo, el 85,8% de personas encuestadas en un estudio de CEPAL para el año 2015 indican que el apoyo debería ser mediante programas que ayuden al desarrollo de habilidades y capacidades, mas no con ayudas en especie (Abramo, Cecchini, & Morales, 2019).

Otro análisis se refiere a las transferencias no monetarias cuyo enfoque es hacia la educación, alojamiento, salud, transporte público, etc. y que va más allá de una ayuda de dinero. Los resultados de un estudio de caso sobre la aplicación de este tipo de transferencias en Reino Unido e Irlanda en el ámbito de la educación para el año 2000 indican que hay efectos importantes en la distribución del beneficio económico. Además, la conclusión general resalta que estas transacciones, hacia cualquier sector, crea una redistribución “pro-poor” (Callan & Keane, 2009).

Teoría del método: algoritmo k-medias

Las técnicas estadísticas con respecto al manejo de datos tienen diversos fines como presentar resultados, resumir información, etc. De hecho la minería de datos se basa en algo similar, “puede crear modelos predictivos, o clasificar elementos, o identificar diferentes grupos de casos dentro de los datos” (Attewell, Monaghan, & Kwong, 2015). Lo cierto es que a pesar de guiarse por fines comunes, cada área usa distintos métodos para trabajar con datos.

Visto de esta forma, la minería de datos se centra además, en trabajar en tareas con fines predictivos y descriptivos. El primero, como se indica, predice una variable con la información de otras variables, entendiendo que busca un resultado matemático o numérico sobre el conjunto de datos. Por otra parte, el fin descriptivo busca “derivar patrones (correlaciones, tendencias, clústeres, trayectorias y anomalías) que resuman las relaciones subyacentes en los datos” (Tan, Steinbach, & Kumar, 2005) en donde reluce la intención de manejar los datos desde un análisis sobre las características de los mismos.

La minería de datos contiene, entre varios, al análisis de datos multivariante, que “comprende el estudio estadístico de diferentes variables medidas en elementos de una población” (Peña, 2002) que como sugiere evalúa conjuntos de datos que van más allá de contener dos variables, algo más cercano a la realidad

Peña (2002) también indica que los objetivos del análisis de datos multivariante son:

- Resumir los datos mediante un pequeño conjunto de nuevas variables, construidas como transformaciones de las originales, con la mínima pérdida de información;
- Encontrar grupos en los datos, si existen;
- Clasificar nuevas observaciones en grupos definidos;
- Relacionar dos conjuntos de variables.

Se trata de crear relaciones entre varios conjuntos de datos con características similares. Como por ejemplo, en el campo empresarial o financiero, estas herramientas permiten clasificar a clientes según condiciones económicas, sociales, territoriales, etc., con el objetivo de ya sea entregar un préstamo o satisfacer segmentos del mercado. Asimismo, en el campo de la documentación ayuda a “clasificar revistas por sus artículos y construir indicadores bibliométricos” (Peña, 2002) de esta manera un análisis preciso puede mejorar una gestión y conducir a una acción.

Dado que la disertación se enfoca en aplicar una técnica que permita el generar grupos del total de los estudiantes de la PUCE-Q, se analiza el contexto teórico de los métodos de clasificación. Un amplio rango de métodos existe para el análisis de similitudes entre grupos. Siendo, entre varios, los más importantes:

- Análisis de correspondencias;
- Perceptron;
- Árbol de decisión;
- Análisis de conglomerados;
- Análisis de componentes principales;
- Máquinas de vectores de soporte.

De entre estos, el análisis de conglomerados se centra en solucionar tres tipos de problemas. El primero de ellos la partición de datos, en donde se presume que los datos “son heterogéneos y se desea dividirlos en un número de grupos prefijado” (Peña, 2002). Seguido de la construcción de jerarquías que busca clasificar según similitud con base en una jerarquía. Para terminar con la clasificación de variables en donde “en problemas con muchas variables es interesante hacer un estudio exploratorio inicial para dividir las variables en grupos” (pág. 220).

En resumen, lo descrito anteriormente muestra la importancia de que la técnica busca clasificar un conjunto de datos o de variables dado que es evidente las similitudes que existen en el mismo. De la misma manera, también hace referencia a que la distribución de elementos se puede hacer de manera jerárquica ya que en estos casos es importante respetar un orden o la relevancia de una característica.

Un clúster por lo general “busca encontrar grupos de observaciones estrechamente relacionadas tal que, las observaciones que pertenecen al mismo clúster son más similares entre sí” (Tan, Steinbach, & Kumar, 2005). Es decir, estudia las características de un grupo de datos desde una perspectiva agrupadora. No obstante, a pesar de que su fin más claro se refiere a una clasificación en donde se encasillan datos en clústeres en el sentido de una distribución técnica, también se considera que la definición más adecuada dependerá de los resultados que aspire a obtener el investigador y a las características del conjunto de datos.

En tal sentido, se puede decir que el análisis de conglomerados “divide los datos en grupos que son significativos, útiles o ambos” (Tan, Steinbach, & Kumar, 2005). Cuando se agrupa con el enfoque significativo, la distribución “captura la estructura natural de los datos” (pág. 487), mientras que si se busca crear clústeres con utilidad “se tiene otros propósitos, como

resumir el conjunto de datos” (pág. 487). Resulta claro que se puede usar la técnica ya sea para clasificar o para presentar información sobre las características del conjunto de datos.

Tipos de clústeres

Según el método de clasificación, los más relevantes son el clúster particional y el clúster jerárquico. El primero se refiere a “una división del conjunto de datos en subconjuntos no superpuestos, de modo que cada elemento está exactamente en un subconjunto” (Tan, Steinbach, & Kumar, 2005) enfatizando que el dato solo puede pertenecer a un solo clúster. El método jerárquico se refiere a un “conjunto de clústeres anidados que están organizados como un árbol” (pág. 492) indicando que existe una ramificación de clústeres, en donde una agrupación es parte de otra que la contiene, respetando así un orden o la relevancia de un clúster sobre otro.

Se puede adicionar que se clasifica según la asignación de datos. Se trata de clústeres exclusivos ya que la observación se asigna a un solo clúster, siendo esto lo contrario a un clúster superpuesto que indica que un elemento puede estar de manera simultánea en más de un grupo. En la misma línea, también está el clúster difuso, en donde “cada objeto pertenece a un clúster con un ponderación de entre 0 y 1” (Tan, Steinbach, & Kumar, 2005) infiriendo que, en base a características, el dato no indica de manera clara su relación a una sola agrupación.

El siguiente punto trata sobre la visión del análisis de conglomerados como un método de clasificación no supervisada, esto refiriéndose a que realiza una distribución automática y aleatoria. Por otra parte, la clasificación supervisada se refiere a “indicar que conocemos una muestra de elementos bien clasificados que sirve de pauta o modelo para la clasificación de las [...] observaciones” (Peña, 2002), y la diferencia radica en que en esta última, se conoce a detalle los criterios del conjunto de datos y el posible número de clústeres.

K-medias

El análisis de conglomerados se puede realizar con diversas herramientas, entre ellas, los algoritmos k-medias, k-medoide, k-modas, DBSCAN, etc. Para el propósito de la investigación se trabajara con el algoritmo k-medias, el cual es “una técnica de agrupamiento particional basada en prototipos que intenta encontrar un número de grupos (k) especificado por el usuario, que están representados por sus centroides” (Tan, Steinbach, & Kumar, 2005). Se refiere a lo descrito en la clasificación de clúster particional, en donde cada elemento va a pertenecer a un solo centro, sin superposición.

Según Peña (2002) el algoritmo procede de la siguiente manera:

1. Seleccionar G puntos como centros de los grupos iniciales;
2. Calcular las distancias euclídeas de cada elemento a los centros de los G grupos, y asignar cada elemento al grupo de cuyo centro esté más próximo;
3. Definir un criterio de optimalidad y comprobar si reasignando alguno de los elementos mejora el criterio;
4. Si no es posible mejorar el criterio de optimalidad, terminar el proceso.

Ahora es preciso detallar cada punto y sus implicaciones. Sobre el primero, la selección de G o K puntos, dependiendo de la notación, puede basarse en diversos criterios. Muchas de las veces el número de grupos ya viene dado por el conjunto de datos, pero caso contrario se puede referir al uso de la prueba F. Este test también llamado de reducción de variabilidad funciona:

Comparando la SCDG¹ con G grupos con la de G + 1, y calculando la reducción relativa de variabilidad al aumentar un grupo adicional [...] y compara la disminución de variabilidad al aumentar un grupo con la varianza promedio. (Peña, 2002)

Esto indica que se aumenta un grupo en la clasificación si es que el resultado de la prueba F sobrepasa la varianza del conjunto de datos. De esta manera, se tiene un estadístico de base que permite justificar el aumento de grupos, si es que no se conoce el número de los mismos previamente.

El segundo punto del algoritmo se refiere al criterio de separación de los grupos. Pues dado que k-medias busca similitud intragrupal y simultáneamente diferencia entre grupos, la pauta que guía este proceso es el cálculo de una distancia entre puntos, y en este caso de clúster a clúster. La distancia euclídea se define como la “medida de proximidad que cuantifica la noción del más cercano para los datos específicos considerados” (Tan, Steinbach, & Kumar, 2005), es decir, esta medida se asegura que cada elemento del conjunto de datos se clasifique en el centroide con observaciones similares.

Con respecto al tercer punto sobre el criterio de optimalidad, es importante recalcar que todos estos métodos tienen la característica de ser no paramétricos. Esto implica que no se rige a los test estadísticos de *p-value* o a intervalos de confianza que demuestren que una clasificación es mejor que otra cuando son comparados. No obstante, algunos de ellos si tienen medidas de error, calculados a partir de su aplicación.

En este caso, se puede medir la calidad del clúster cuando “se prefiere al (modelo) con la menor suma de cuadrados dentro de los grupos, ya que esto significa que los prototipos (centroides) del clúster son una mejor representación de los puntos dentro del clúster” (Tan, Steinbach, & Kumar, 2005). En efecto, el minimizar la distancia entre los puntos del clúster es un indicativo de una mejor clasificación en relación a los otros clústeres.

Asimismo, otra opción para determinar que la clasificación es más cercana a un criterio óptimo es el ratio de desemejanza que se define como “la relación entre la distancia entre grupos y la distancia dentro del clúster” (Attewell, Monaghan, & Kwong, 2015). Al contrario que la medida de SCDG, aquí se busca maximizar este ratio para que los elementos caigan en un grupo lo más similar posible.

En el programa estadístico R la medida de optimización es $\text{between_SS} / \text{total_SS}$ y así como R^2 , utilizado en regresiones econométricas, “indica el porcentaje de varianza explicada por el modelo respecto al total de varianza observada” (Amat, 2017). El primer componente, between_SS indica la suma de las distancias cuadradas del centroide a la media global de la muestra, mientras que total_SS se refiere a la suma de las distancias cuadradas de la observación a la media global de la muestra. Entonces el indicador se refiere a que tan bien la muestra se divide en los clústeres designados.

Previo al paso final, Attewell, Monaghan, & Kwong; se refieren a ciertas consideraciones sobre el proceso que realiza el algoritmo con respecto a la media:

¹ Suma de cuadrados dentro de los grupos.

En este paso, el algoritmo k-medias encuentra la media del centros de cada una de las nubes de puntos (es muy poco probable que la media sea el punto escogido inicialmente), y establece a estos puntos como los nuevos centroides. Repite el paso anterior- calcula distancias, asigna casos a los centroides, encontrando puntos de media y cambiando centroides- una y otra vez hasta que el programa converga a una solución estable. (Attewell, Monaghan, & Kwong, 2015)

Lo analizado en esta sección sintetiza el proceso por el que pasa un conjunto de datos bajo el cálculo del algoritmo para obtener una clasificación lo más cercana a lo óptimo, dados los criterios enunciados anteriormente.

Capítulo I: Sistema de categorización de la PUCE-Q

Introducción

Parte de la visión legal de garantizar educación terciaria, el enfoque se asegura de que los procesos universitarios se centren también en parámetros de excelencia. Por lo mismo, la Pontificia Universidad Católica del Ecuador (PUCE) guía su proyecto institucional en el modelo educativo de la PUCE, que a grandes rasgos constituye, la base de la gestión académica de la universidad. Este modelo busca desarrollar cuatro principios; “la centralidad del estudiante en el proceso de enseñanza y aprendizaje, el humanismo cristiano, la pedagogía ignaciana y la integración del saber” (PUCE, 2017). En lo que concierne a la centralidad del estudiante en el proceso de enseñanza y aprendizaje, se debe recalcar la relación con la importancia de cumplir el objetivo de acceso a la educación. Esto último mediante la organización de un sistema que permite clasificar a los estudiantes según sus características económicas y sociales.

Para el cumplimiento con el principio antes mencionado, la PUCE cuenta con el Sistema de Beca Socioeconómica (SBS) que sigue parámetros de clasificación social y económica. La clasificación que realiza se basa en las características que presente el estudiante y su entorno resultando en la asignación de una categoría. El SBS puede asignar una categoría de las cinco que compone, a cada estudiante, las mismas pueden ser A, B, C, D y E-F². Siendo A la que mayor descuento recibe y E-F la que no recibe ningún descuento.

El principal objetivo del SBS es descontar del total de matrícula y créditos, un monto de acuerdo a la clasificación asignada. Lo mismo permite potenciar las oportunidades del estudiante para aprovechar los niveles de estudio, sin ninguna clase de interferencia que limite su educación.

1.1. Descripción de beneficios económicos y categorías

La Pontificia Universidad Católica del Ecuador, siguiendo con los parámetros anteriores, “cuenta con un importante Programa de Beneficios Económicos para sus estudiantes, los mismos que se otorgan bajo los criterios de: equidad, inclusión, excelencia y pertenencia” (PUCE, s.f.). Este programa ayuda a la población estudiantil, clasificándola con el fin de equilibrar las condiciones en el pago de sus estudios.

La PUCE- Q define a una beca como “la subvención total o parcial que la universidad otorga a sus estudiantes” (PUCE, 2017), con el objetivo de consecución de sus estudios. Por lo general, las becas otorgadas se basan en características del entorno del estudiante, así como ayudan a reconocer y premiar habilidades y/o capacidades académicas.

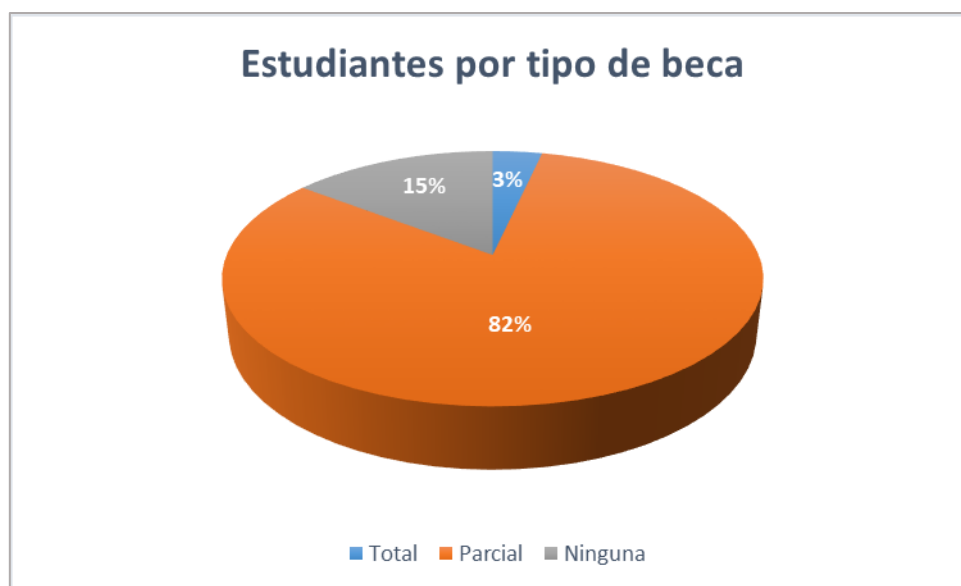
Por lo mismo, la PUCE-Q en su Reglamento General de Estudiantes (2017) dispone de cinco categorías de becas:

² La categoría E-F se presenta así debido a la transición de Sistema de Pensión Diferenciada a Sistema de Beca Socioeconómica, en donde estas dos categorías convergieron.

- Becas socio-económicas: son aquellas que la universidad concede a estudiantes de escasos recursos económicos, en función de su ubicación en una de las categorías socio-económicas reconocidas en la universidad;
- Becas inclusión: son aquellas que la universidad concede a estudiantes susceptibles de ser excluidos por su pertenencia a grupos históricamente excluidos, por su género, situación de discapacidad o pertenencia a grupos sociales con derechos específicos;
- Becas académicas: son aquellas becas que la universidad puede otorgar a todos los estudiantes regulares que cumplen semestralmente con el promedio ponderado de notas que fije para el efecto cada una de las unidades académicas [...];
- Becas de desarrollo personal: son aquellas becas que la universidad puede otorgar en apoyo de la formación integral del estudiante. Son de tres tipos: relacionados con cultura, arte y deporte [...];
- Becas institucionales: son aquellas becas que la universidad puede otorgar en razón de sus principios o prioridades como institución, y que no entran en ninguna de las tres categorías anteriores, dependiendo de la disponibilidad presupuestaria.

Una manera diferente de evidenciarlo es presentando información sobre los estudiantes que reciben una beca total o parcial:

Gráfico 1: Porcentaje de estudiantes de la PUCE-Q por tipo de beca total o parcial (2018).



Fuente: Dirección General de Estudiantes.

Elaborado por: Dayana S.

La mayoría de los estudiantes, el 82%, recibe una beca parcial. Es decir, algún beneficio como los descritos anteriormente, que ayuda al estudiante en el pago de sus estudios. El 3% recibe una beca total y un 15% no recibe una beca y paga el valor total de su matrícula y sus créditos.

Este conjunto de beneficios económicos permite obtener de acuerdo con sus méritos académicos; becas, créditos y otras formas de apoyo económico que, según el artículo 5 de la LOES, le “garantice igualdad de oportunidades en el proceso de formación de educación superior” (LOES, 2018). Como fue desarrollado en la fundamentación teórica, esto permite el acceso, desarrollo y consecución de los estudios de educación superior del estudiante.

Siendo uno de los beneficios económicos más importantes, la beca socioeconómica. Por la cual todos los estudiantes admitidos pasan por un análisis a través de un sistema informatizado y dependiendo de las condiciones económicas de su grupo familiar, se le otorga una categoría con un descuento que va desde el “25% al 55%, sobre los aranceles semestrales” (PUCE, s.f.) (véase el gráfico 2).

Una vez que el estudiante ingresa en la universidad en calidad de matriculado se somete al proceso de asignación de beca socioeconómica y es clasificado de manera obligatoria en una de las categorías del SBS. Siempre preservando el principio de igualdad de oportunidades para los estudiantes. Esta primera beca que brinda la universidad a sus estudiantes, sigue un proceso de revisión cada cuatro semestres, cuyo fin es analizar potenciales variaciones socioeconómicas que cada hogar pueda enfrentar (PUCE, 2017).

Los detalles de los límites que comprende la división de cada categoría en valores numéricos o porcentuales indican que reciben un descuento por categoría que se puede visualizar de la siguiente manera:

Gráfico 2: Visualización de categorías según el descuento recibido (2018).

Categoría A	• hasta el 55% de descuento.
Categoría B	• hasta el 45% de descuento.
Categoría C	• hasta el 35% de descuento.
Categoría D	• hasta el 25% de descuento.
Categoría E	• ningún descuento.

Fuente: entrevista a Dirección de Beneficios Económicos.

Elaborado por: Dayana S.

El gráfico 2 indica que cada categoría está ubicada en un nivel que representa el monto de descuento que recibe. Es decir, la categoría A recibe el mayor monto de descuento, y así sucesivamente con las demás categorías que conforman el gráfico.

La metodología para la clasificación de los estudiantes en las categorías se basa en rangos de la puntuación obtenida de los componentes del modelo de categorización que se detallará en el siguiente apartado. De manera general, se resume:

La lógica de la puntuación radica en asignar un puntaje más alto al individuo con mayor vulnerabilidad. El identificador 5 (Categoría 1= A) corresponderá a aquel individuo de la categoría más baja mientras que el identificador 1 (Categoría 5=E y F) corresponderá a la categoría más alta (Muñoz Arias, 2018).

Lo anterior indica que la clasificación se basa, entre otros, en el componente de vulnerabilidad. A pesar de que el análisis comprende una dimensión económica, el proceso prioriza las condiciones de vulnerabilidad para asignar la categoría. Se indica que “este proceso de estratificación, irá priorizando la vulnerabilidad para cada 20% de los estudiantes evaluados” (Muñoz Arias, 2018) (véase el cuadro 1):

Cuadro 1: *Asignación de categoría según rango de puntuación del modelo.*

Categoría	Distribución por cada 20%
Categoría 1	Rango puntaje Quintil 5 muestra
Categoría 2	Rango puntaje Quintil 4 muestra
Categoría 3	Rango puntaje Quintil 3 muestra
Categoría 4	Rango puntaje Quintil 2 muestra
Categoría 5	Rango puntaje Quintil 1 muestra

Fuente y elaboración: Dirección de Beneficios Económicos (2018).

1.2. Descripción general del modelo actual de categorización

Dado el carácter fiscomisional de la PUCE, la Ley de Educación Superior en el artículo 90 indica “cobros de aranceles diferenciados en las instituciones de educación superior particulares” (LOES, 2018). Dado además que la PUCE es fiscomisional, el Sistema de Beca Socioeconómica permite que todos los estudiantes paguen el costo semestral de los créditos académicos y matrícula en forma proporcional a su situación económica, personal y familiar.

Lo anterior, se enlaza con la pedagogía ignaciana, en la cual la PUCE basa parte de su accionar en aspectos de docencia, investigación, etc. Uno de los cinco pasos de esta metodología se refiere a la contextualización, en donde se resalta la importancia de evaluar las circunstancias y los protagonistas de una situación. Por lo mismo, se crea un entendimiento sobre el entorno del estudiante y su familia (PUCE, 2017).

Los lineamientos del Sistema de Beca Socioeconómica son aplicados en todas las sedes de la PUCE a nivel nacional. De hecho, los antecedentes, objetivos y procesos son emulados en los manuales de procedimientos de las cinco sedes, en concordancia con lo establecido por la matriz Quito. Por lo mismo, se toma información de los registros encontrados en la sede Ibarra, que explica el accionar del SBS.

Básicamente, una de las finalidades es que:

Este sistema permite fijar el costo semestral de los créditos académicos y matrícula del estudiante y la universidad mantiene abierta la posibilidad de permitir el ingreso al estudiante académicamente calificado, al margen de sus posibilidades económicas. (Comisión de becas PUCE-SI, 2012)

Es decir, el sistema intenta garantizar educación en base a las capacidades académicas del estudiante, sin que las condiciones de pago supongan una carga durante su periodo de estudios. De la misma manera, el sistema se empareja con el objetivo de conseguir una

distribución y uso de los recursos que compagina con una gestión solidaria por parte de la universidad (Comisión de becas PUCE-SI, 2012).

El denominado hasta 2016 Sistema de Pensión Diferenciada (S.P.D.), se basa en varias políticas generales, entre ellas las más importantes (Comisión de becas PUCE-SI, 2012):

- Planificar periódicamente el proceso de ubicación o reubicación de los estudiantes en el S.P.D.;
- Establecer pautas e innovaciones que permitan la eficiente ubicación de los estudiantes en las distintas categorías del S.P.D.;

La existencia de un proceso como el presentado anteriormente que conste de pautas generales, permite un ordenamiento y continuación del proceso que no debería presentar situaciones irregulares y que logre regularlas si es que se producen. Asimismo, estas políticas involucran tanto los procesos, como los responsables para la consecución del Sistema de Beca Socioeconómica.

Para la clasificación de los estudiantes, la PUCE (2018) recurre a un método que utilice la documentación requerida y presentada además de considerar entrevistas personales. Los documentos por considerar son:

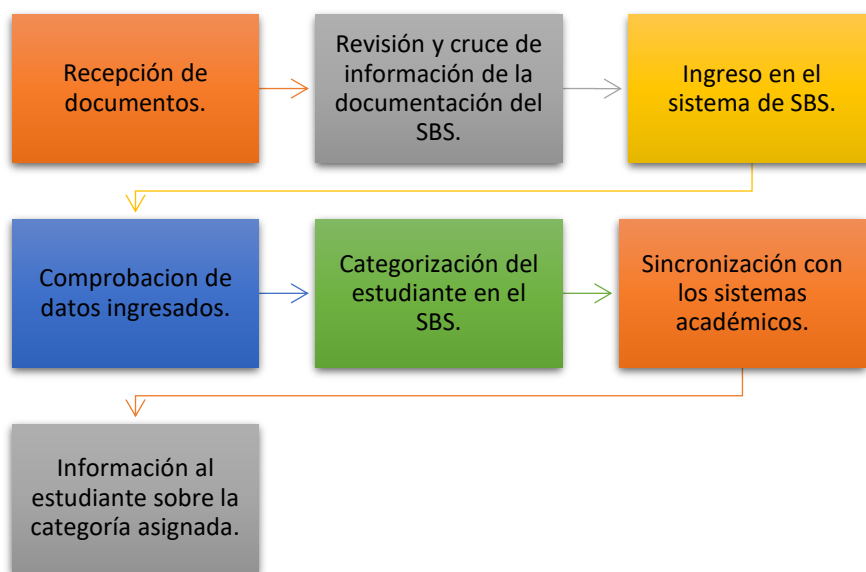
1. Para estudiantes nuevos, comprobante del último pago de pensión de colegio, Acta de Grado o Título de Bachiller.
2. Para estudiantes en proceso de re-categorización, comprobante de pago o factura del último semestre académico.
3. En caso de estudiantes casados, copia de la cédula que refleje su estado civil.
4. En caso de que el estudiante tenga hijos, partida de nacimiento del o los hijos.
5. En caso de tener RUC, último formulario de declaración de Impuesto a la Renta e IVA (aplica para estudiantes y representantes).
6. Tabla de amortización crédito hipotecario (una sola vivienda).
7. Estados de cuentas bancarias y tarjetas de crédito de los últimos 3 meses, certificadas por la entidad financiera.
8. Si el estudiante (u otro miembro de su familia) tiene una discapacidad, adjuntar el carné de discapacidad.
9. En caso de enfermedades catastróficas o degenerativas del estudiante, un certificado médico actualizado.
10. Croquis de la dirección actual del estudiante.

La documentación que se presenta varía según otras características del estudiante como por ejemplo, si los estudiantes viven independientemente de sus padres, estén casados, presentan discapacidades, son refugiados o extranjeros. Asimismo, si los padres de los estudiantes presentan situaciones en las que: están jubilados, divorciados o han fallecido, se deberá demostrar la condición para recibir un trato adaptado a la situación en la asignación de su categoría.

La Dirección General de Estudiantes (DGE) y la Dirección de Beneficios Económicos (DBE) son los actores principales del proceso y orientan a los estudiantes hacia el proceso del SBS. Informan al estudiante sobre el costo de aranceles y matrícula según carrera y número de créditos. Además, se encargan de la entrega del formulario del SBS junto a los requisitos para la categorización.

La Dirección de Beneficios Económicos, por su parte se encarga de la gestión de la información para designar la categoría del estudiante:

Gráfico 3: Proceso de uso de información para asignación de categoría SBS.



Fuente: Comisión de becas PUCE- SI (2012).

Elaborado por: Dayana S.

El diagrama anterior demuestra que dentro del proceso de manejo de información de los documentos proporcionados por los estudiantes, existen actividades direccionadas a controlar y comprobar la veracidad de lo entregado. De igual manera, se muestra que la información queda en los sistemas informáticos de la universidad como parte de respaldos del proceso así como para la recategorización que procede más adelante.

Cabe destacar que este trabajo de investigación no se enfocará en el análisis del proceso de recategorización, que se supone utiliza parámetros de evaluación diferentes y no conocidos por el investigador.

1.2.1. Metodología del modelo

Cabe recalcar que la información sobre la metodología del modelo no fue contrastada con entrevistas a funcionarios de la Dirección General de Estudiantes. Lo presentado en este apartado corresponde al documento “Proyecto Beca Socioeconómica” del año 2018, proporcionado por la misma unidad administrativa. Por lo mismo, se reconoce las limitaciones e implicaciones de lo indicado.

El modelo se basa en el uso de variables cualitativas y cuantitativas, que recojan las condiciones socioeconómicas del estudiante y su entorno. Por lo mismo, tiene una base conceptual en Alkire et al. (2011) que “logra integrar una visión multidimensional de la pobreza desde el acceso a servicios públicos y sociales”. En efecto, se busca no solo medir la

capacidad de ingresos de la familia, sino también equiparar esto con un análisis de condiciones básicas de vida (Muñoz Arias, 2018).

Es conveniente indicar que al igual que la vulnerabilidad, también se incluye:

[...] elementos del componente “Protección al Ciclo de Vida” con el propósito de estructurar un mecanismo de compensación que fomente la equidad social desde la perspectiva de una evaluación de las condiciones que impiden a una persona acceder a un trabajo formal. (SENPLADES, 2014 citado en Dirección de Beneficios Económicos, 2018)

Coincidiendo en equilibrar la situación de personas con condiciones económicas desfavorables bajo el marco de guiar el proceso con una consigna social, en este caso la equidad. La misma se define con un concepto de justicia intertemporal. En la cual, se buscaría llegar a todos, sin embargo, su foco más realista es llegar a la mayoría, a la que se puede asignar la misma cantidad de recursos en diferentes periodos de tiempo (Mokate, 1999).

En base a lo analizado el modelo de categorización “busca evaluar las condiciones socioeconómicas de las familias de los estudiantes desde 3 dimensiones: Ingresos, Gastos y otras condiciones de vulnerabilidad” (Muñoz Arias, 2018) incorporando lo que se conoce como análisis socioeconómico.

Los elementos que componen cada dimensión son los siguientes:

Cuadro 2: Componentes de cada dimensión del modelo de categorización de la PUCE-Q.

Gastos (acceso los bienes superiores)	Ingresos	Otras condiciones de vulnerabilidad
Gasto Salud	Ingreso total del hogar per cápita (por concepto de rentas, arriendos, pólizas, préstamos a terceros, etc.)	Discapacidad
Gasto Educación		Tipo de Ocupación: ocupación, desempleo. Endeudamiento corriente.
Gasto Vivienda		Condiciones de Habitación y vivienda: tenencia de la vivienda, acceso a servicios públicos (agua, luz, alcantarillado), bienes y servicios (internet), ubicación geográfica.
Gasto Alimentación		Educación: tipo de colegio, crédito educativo.

Fuente: Dirección de Beneficios Económicos (2018).

Elaborado por: Dayana S.

Por supuesto, los datos que arrojan estos componentes no están en la misma escala o formato. Asimismo, en el caso de la dimensión de ingresos, pueden aparecer valores atípicos que de cierta manera distorsionen el análisis y la aplicación del modelo. Por esto, cada dimensión tiene un proceso de estandarización que permite evadir estas situaciones propias de la información.

Dimensión de gastos o acceso a bienes superiores

Básicamente, indica que “a partir de la proporción del gasto se logra determinar el nivel de acceso a las necesidades básicas desde la capacidad del consumo de los bienes superiores por parte de los hogares”. De hecho este método de dimensionar el acceso a través del gasto familiar es una forma de conocer una parte importante de la estructura financiera familiar del estudiante (Muñoz Arias, 2018).

La estandarización de la Dirección de Beneficios Económicos (2018) para todos los componentes de gasto sigue los siguientes pasos. (Ver Anexo n.º 1):

1. Proporcionar: se determina el gasto total y se divide para el ingreso total;
2. Normalizar: para tratar los datos atípicos y se aplica que a “los valores menores al percentil 10 se codifica con un valor igual a 0 y los valores mayores al percentil mencionado se les ubica con un valor igual a la proporción del gasto” (Muñoz Arias, 2018);
3. Recodificar: transformar la variable a tipo categórica en función “a los deciles generados a nivel de Quito para la variable en cuestión, aplicando los datos de ENINGUR (INEC, 2012) donde “se divide [respectiva], para la variable ‘ing_cor_tot’ que corresponde al Ingreso Corriente del Hogar”;
4. Estandarizar: “los rangos de proporción de gasto de cada variable en rangos entre 10 y 100”.

Dimensión de ingresos

Como se indicó, esta dimensión incluye todos los montos monetarios de cualquier concepto, es decir, el dinero recibido por ejercicio profesional, laboral, negocio más otros ingresos (véase el cuadro 2). Por tanto, se puede señalar que “se recogerá las brechas de desigualdad y distribución del ingreso medio de los estratos sociales” (Muñoz Arias, 2018).

Y se sigue un proceso similar al de la estandarización de gastos, con las siguientes consideraciones (Muñoz Arias, 2018):

1. Proporcionar: se determina el ingreso total y se divide para el número de perceptores del hogar.
2. Normalizar: ordenando de forma ascendente el cálculo del paso 1.
3. Recodificar y estandarizar: “el decil más bajo tendrá una puntuación de 100 y el decil más alto de 10”.

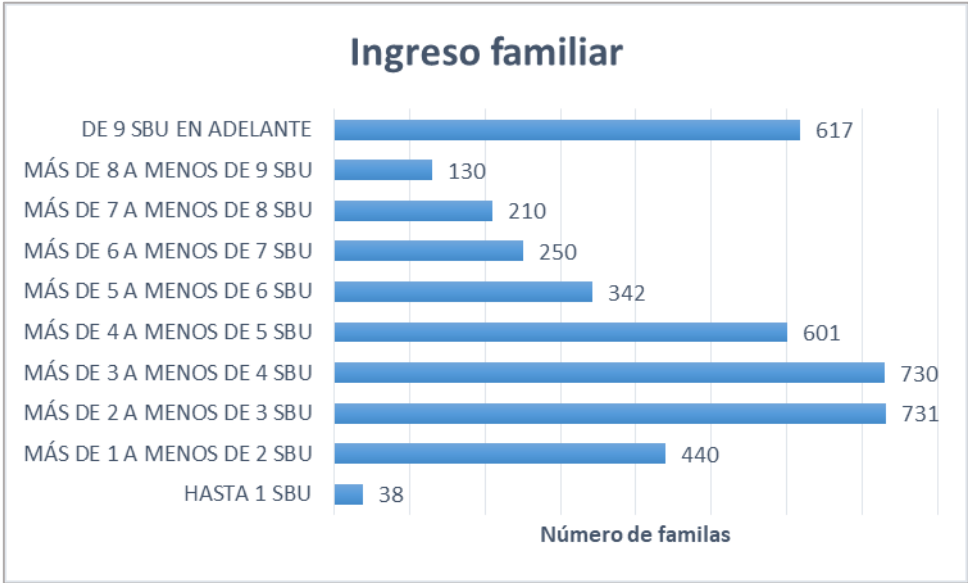
Adicionalmente, se toman en cuenta otras consideraciones sobre posibles fuentes de ingresos por ciertos conceptos, por ejemplo, las personas que cuenten con un negocio propio:

Se evaluará el volumen de ventas generado, cuya validación referirá a documentos como el Impuesto a la Renta, Declaración al valor Agregado, Régimen impositivo simplificado. Para ello se ha establecido un rango de ventas que calcula un margen de utilidad del 10% por tipo de negocio y en función al ingreso generado se divide para 4 miembros del hogar (número de miembros promedio por familia). (Muñoz Arias, 2018).

Como resultado este análisis arroja seis rangos de volumen de ventas basado en el cálculo de ingreso del paso 1. Se determina que para los dos primeros niveles la información ya se puede aplicar el proceso de SBS. Con relación a los cuatro niveles restantes, donde el volumen de ventas supera los \$20.000,00, se asigna automáticamente a la categoría 5 (ver Anexo n.º 2) (Muñoz Arias, 2018).

Los datos de la Dirección General de Estudiantes arroja la siguiente distribución sobre los ingresos familiares de los estudiantes de la PUCE-Q:

Gráfico 4: Ingreso familiar de los estudiantes de la PUCE-Q (2018).



Fuente: Dirección General de Estudiantes.
Elaborado por: Dayana S.

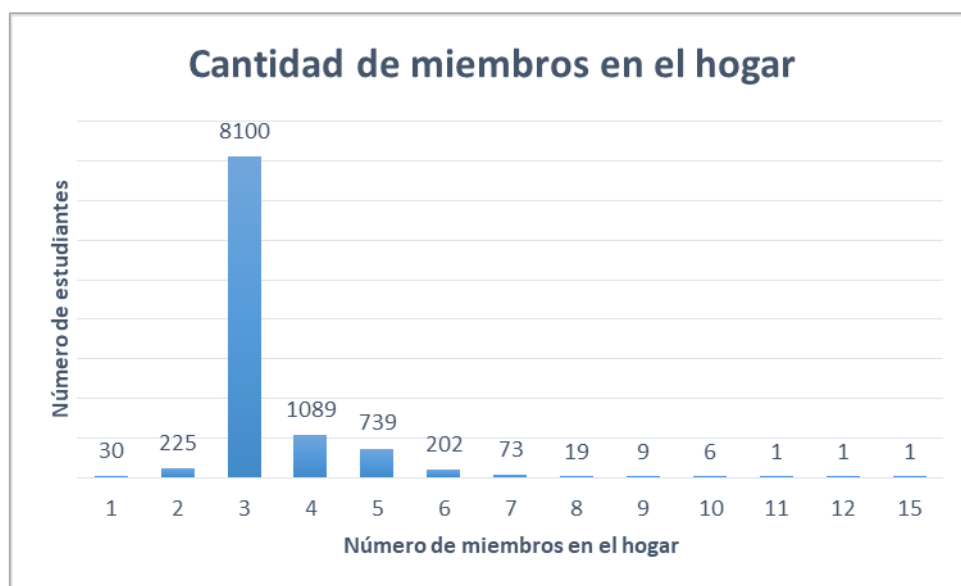
En el gráfico anterior se suprimió 6.407 registros de las personas que indicaron “No registra” pues distorsionaba el análisis. El 17% indica que su ingreso es de “más de 3 a menos de 4 SBU³”. Considerando que en 2018 el SBU se fijó en \$386,00 dólares, se diría que la mayoría de las familias reporta un ingreso de aproximadamente, entre \$1.158,00 y \$1.544,00 dólares (Ministerio del Trabajo, 2018).

Es interesante observar que 38 familias perciben ingresos de hasta 1 SBU; \$386,00 dólares. Por otra parte, se ve que el 15% reporta un ingreso de 9 SBU en adelante, es decir, más de \$3.474,00 dólares.

Sobre los resultados, también se pueden agregar un análisis de las características principales de la población de estudiantes que influyen en una clasificación socioeconómica. Por ejemplo, la variable de “cantidad de miembros en el hogar”:

³ Salario Básico Unificado.

Gráfico 5: Cantidad de miembros en el hogar de los estudiantes de la PUCE-Q (2018).



Fuente: Dirección General de Estudiantes.

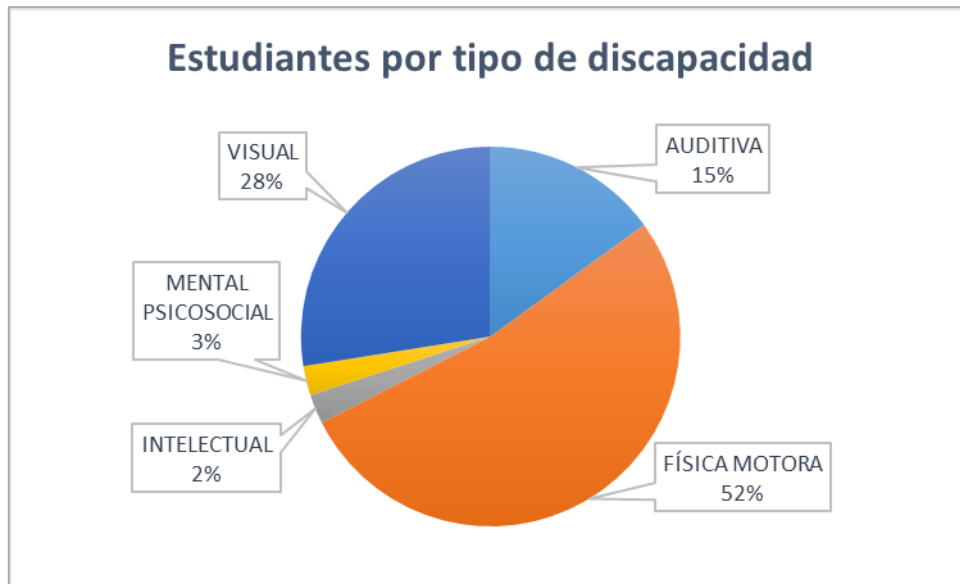
Elaborado por: Dayana S.

Se evidencia que el 77% de los estudiantes indica que existen tres miembros en su hogar. Además se observa que la mayoría de los datos se agrupa en el rango de 1 a 7 miembros por hogar. Son muy pocos los casos en donde se registran más de 10 personas. Esto brinda cierta perspectiva al cálculo de ingreso per cápita que se realiza en esta dimensión.

Dimensión de otras condiciones de vulnerabilidad

En este apartado se considera características que implican o están asociadas a niveles de vulnerabilidad dentro del hogar. Con relación al análisis, esto se refiere a discapacidad del estudiante, acceso a servicios básicos, entre otros (véase el cuadro 2). En este caso, la estandarización solo sigue la directriz de asignar 100 si se cumple la característica de vulnerabilidad o de lo contrario se indicará cero (ver Anexo n.º 3) (Muñoz Arias, 2018). Uno de los componentes más importantes de esta dimensión se refiere a si el estudiante tiene algún tipo de discapacidad o una enfermedad catastrófica:

Gráfico 6: Porcentaje de estudiantes con discapacidad de la PUCE-Q por tipo de discapacidad (2018).



Fuente: Dirección General de Estudiantes.

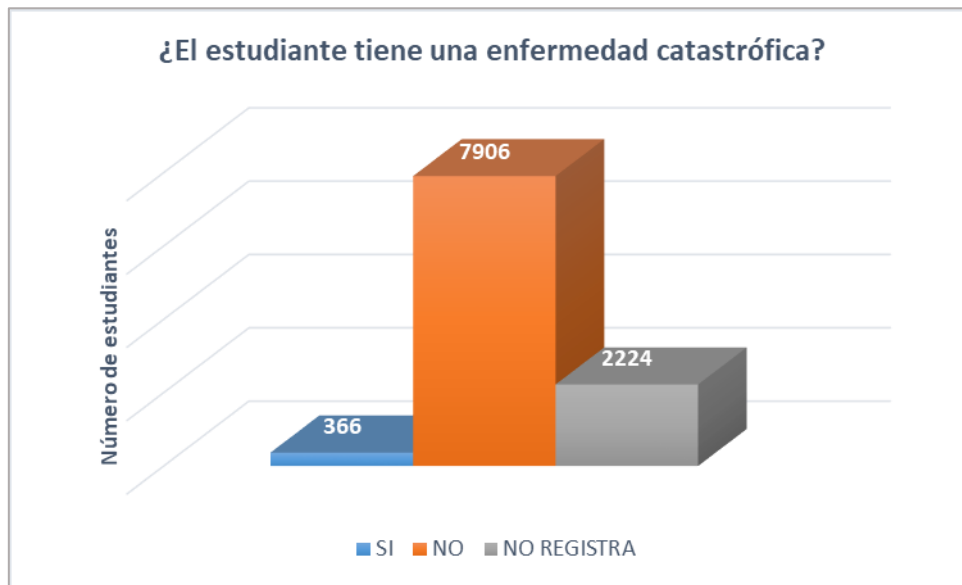
Elaborado por: Dayana S.

Se recalca que de los 40 estudiantes que presentan discapacidad, el 52% presenta una de tipo física motora, la cual restringe la movilidad de determinadas partes del cuerpo. Además, se indica que un 28% de los estudiantes presentan discapacidad visual y el 15% de tipo auditivo. En el primer caso, la discapacidad impide una comprensión del contenido educativo de manera visual. Mientras que el tipo de discapacidad auditiva implica un uso de recursos y técnicas diferentes, como el lenguaje de señas para procurar un aprendizaje adecuado (INEGI, 2020).

Aunque con un porcentaje relativamente pequeño del 3%, el tipo de discapacidad mental psicosocial, se refiere a personas que tienen un diagnóstico relacionado a una enfermedad mental o que han experimentado discriminación, exclusión o estigma. Mientras que el 2% se centra en la discapacidad intelectual que implica una deficiencia en el proceso de aprendizaje de la persona (Organización Mundial de la Salud, 2020).

En la misma línea, se presenta el estadístico sobre estudiantes con enfermedades catastróficas:

Gráfico 7: Estudiantes de la PUCE-Q con enfermedad catastrófica (2018).



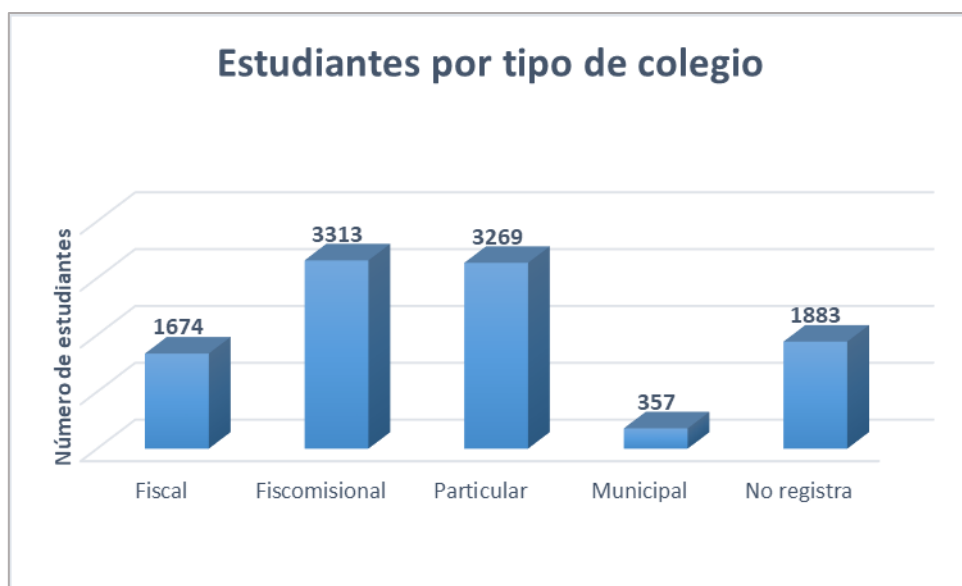
Fuente: Dirección General de Estudiantes.

Elaborado por: Dayana S.

En este caso, el 3% de los estudiantes en este periodo indican tener una enfermedad catastrófica, el 75% no presenta esta condición. Por otra parte, el 22% restante no registra información sobre esta categoría, y con esto se existe una ambigüedad en donde no se sabe si son personas que tienen o no este tipo de enfermedad.

Otra consideración en esta dimensión es el tipo de colegio del estudiante:

Gráfico 8: Estudiantes de la PUCE-Q por tipo de colegio (2018).



Fuente: Dirección General de Estudiantes.

Elaborado por: Dayana S.

Sobre esto, el 32% de estudiantes provienen de colegios fiscomisionales, y el menor porcentaje, 3% estudiaron en colegios municipales. El 18% de observaciones no registran el tipo de colegio, lo que deja la misma conclusión de análisis del gráfico 7, en donde no se sabe si realizaron sus estudios en las opciones de tipos de colegios presentados.

Ponderaciones del modelo

Se vuelve a hacer referencia a la teoría de Alkire (2011), y este caso se le atribuye que el valor del ponderador dependerá del número de dimensiones determinadas refiriéndose a las tres dimensiones detalladas, por lo mismo, a cada dimensión le corresponde un valor de 0.3333. Asimismo, se le asigna un peso proporcional del total, a cada elemento o subdimensión que compone cada dimensión.

Cuadro 3: *Ponderación de las dimensiones y subdimensiones del modelo de categorización de la PUCE-Q.*

Dimensión	Subdimensión	Ponderador
Ingresos (0.33)	Ingreso per cápita	0,3333
Gastos (0.33)	Vivienda	0,0833
	Salud	0,0833
	Alimentación	0,0833
	Educación	0,0833
Otras Características (0.33)	Ubicación Geográfica	0,0417
	Condiciones de propiedad de la vivienda	0,0417
	Ocupación	0,0417
	Servicio: internet	0,0417
	Acceso a Servicios Luz, alcantarillado y agua potable	0,0417
	Discapacidad	0,0417
	Tipo de Colegio	0,0417
	Endeudamiento Corriente	0,0417

Fuente y elaboración: Dirección de Beneficios Económicos (2018).

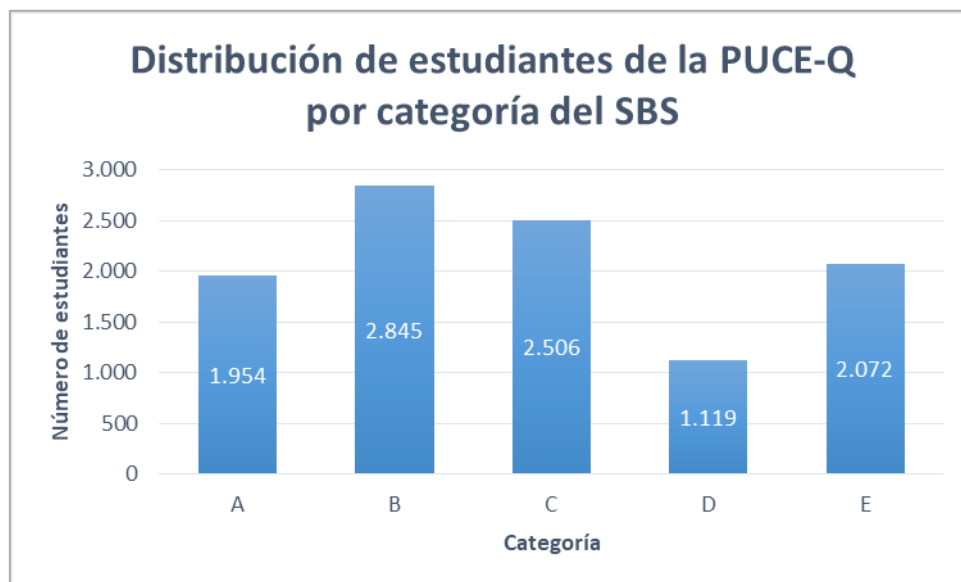
1.3. Análisis de resultados de la distribución de estudiantes según las categorías del Sistema de Beca Socioeconómica 2018

1.3.1. Análisis de resultados 2018

El desarrollo de este apartado será complementario para el análisis del tercer objetivo, en donde se busca comparar los resultados de modelo explicado en la sección anterior con los del modelo desarrollado en el capítulo 2. Los resultados corresponden al año 2018.

Como parte de los resultados de la PUCE sobre la distribución de los estudiantes en categorías se obtiene que, de 10.496 estudiantes:

Gráfico 9: Distribución de estudiantes por categoría – PUCE-Q (2018).



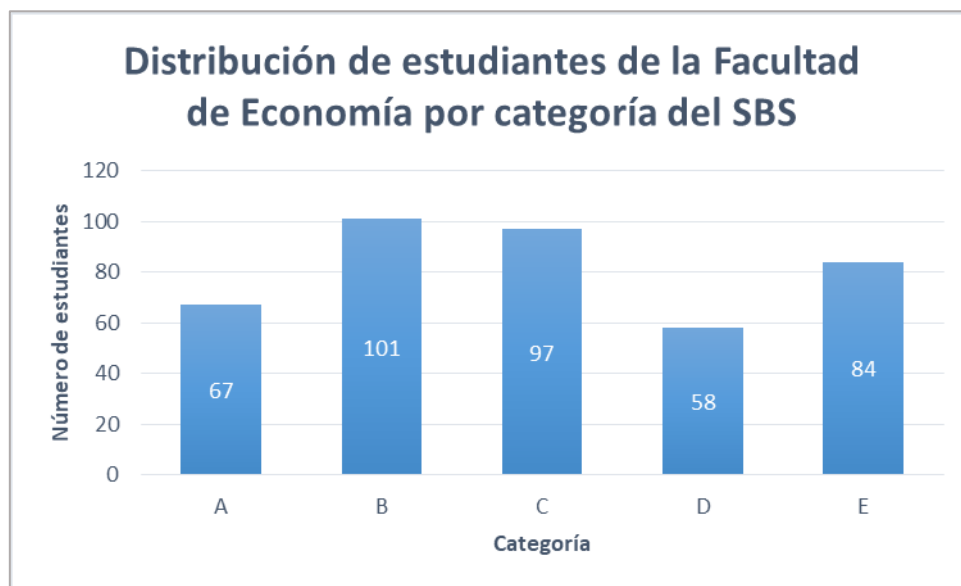
Fuente: Dirección General de Estudiantes.

Elaborado por: Dayana S.

El gráfico destaca que dentro de la distribución de la población estudiantil, la categoría que predomina es la B, asimismo, se observa que la categoría con menos estudiantes es la D. Es decir, la mayoría de estudiantes recibe hasta el 45% de descuento en el pago de sus créditos estudiantiles y, por otra parte, una proporción menor de estudiantes recibe hasta el 25% de estudiantes.

Para un análisis diferente se aplica la misma lógica a la distribución de los estudiantes de la Facultad de Economía. De los 407 estudiantes:

Gráfico 10: Distribución de estudiantes por categoría- Facultad Economía (2018).



Fuente: Dirección General de Estudiantes.

Elaborado por: Dayana S.

Los resultados aplicados en la Facultad de Economía se asemejan a los de la PUCE-Q en general, siguen una tendencia similar en la distribución de estudiantes por categoría. Por ejemplo, la categoría que agrupa la mayor cantidad de estudiantes es la B y por el contrario, la que menos estudiantes agrupa en la D.

Se responde al primer objetivo de la disertación que se refiere a describir el Sistema de Beca Socioeconómica de la PUCE-Q para la asignación de becas de los estudiantes con respecto a su situación económica y entorno familiar. Se procedió a incluir las dimensiones que describen la situación del estudiante a través de detallar los documentos entregados y que el Sistema de Beca Socioeconómica y sus responsables gestionan para la asignación de la categoría. Esto incluye un análisis del estudiante tanto en su posición económica y social, así como el de su entorno familiar o el de sus representantes.

De igual manera, se describió a grandes rasgos el proceso detrás del Sistema de Beca Socioeconómica, que la PUCE- Q adopta como parte de lineamientos sobre acceso a la educación y que impulsan la continuidad de los estudios de los estudiantes, así como un uso más responsable de los recursos que recibe.

La última parte sobre el análisis de los resultados del año 2018 obtenidos con el modelo actual de SBS, reflejan que si bien, la distribución de estudiantes no es perfecta, tampoco es desigual o presenta sesgos extremos. Este análisis será útil para comparar el modelo por construir en el siguiente capítulo con estos resultados y convenir alguna mejora o evidencia un cambio que impacte en beneficio de los estudiantes.

Capítulo II: Modelo alternativo de categorización

2.1. Motivación

El manejo de recursos para la asignación de becas a estudiantes es un tema a tratar con bases sociales y económicas. Por esto, el adecuar un modelo, generalmente cuantitativo, con variables que describen la perspectiva social de un estudiante es una gran tarea y responsabilidad. La construcción de esta clase de modelo siempre buscará clasificar de la mejor manera con el fin de beneficiar al estudiante en cuestión.

De hecho, como se describió en la sección metodología, varios trabajos de investigación tratan con la aplicación de varias técnicas para la consecución de un modelo coherente. En común, la mayoría de modelos coinciden en incluir dimensiones sociales y económicas. No obstante, es interesante observar que el tratamiento de variables y datos difiere según el objetivo por alcanzar.

Es preciso indicar que en estos estudios se utiliza técnicas de minería de datos y de análisis de datos multivariantes. De igual manera, consiguen una clasificación socioeconómica acorde a los objetivos planteados y concuerdan en recomendar actualizaciones del modelo, así como revisión de las variables indicadas. No generan un modelo perfecto pero muestran la posibilidad del uso de varias técnicas estadísticas.

Visto de esta forma, el uso de técnicas de minerías de datos se ha incrementado últimamente en diversos ámbitos. Tal parece que las características de trabajar con conjuntos de lo denominado “big data”, así como la posibilidad de acceder a ellos con un algoritmo disponible en varios programas gratuitos de programación; hacen que sea una opción atrayente para el tratamiento de datos e información.

Sin embargo, a pesar de las ventajas de estas herramientas, es bueno recalcar la relevancia de modelos como el que utiliza la PUCE. Posterior al análisis del capítulo 1 donde se brinda los detalles del modelo, cabe rescatar el enfoque que se hace sobre la dimensión de componentes de vulnerabilidad. Esto busca recoger las características que modelan la realidad del estudiante, que va más allá del monto que puede pagar e incluye un análisis de su entorno en general.

Se quiere con esto indicar que el modelo que se realiza a continuación trata de recoger los aspectos de las dos ópticas. Busca emparejar el uso de la técnica con las características de estudio sobre vulnerabilidad, y esperando que el esfuerzo realizado esté alineado con el objetivo de probar una técnica de clasificación que respete consideraciones socioeconómicas y de la técnica en sí.

2.2. Descripción general del modelo alternativo

2.2.1. Procesamiento de la base de datos

La base de datos de la Dirección General de Estudiantes se refiere la información auto reportada de los estudiantes nuevos que ingresaron en el año 2018, aclarando que no se trabajó con la información proporcionada por el buró. Contiene 2.163 observaciones/registros y 49 variables que reflejan la información general del estudiante y la información de ingresos de dos representantes. No se encontraron observaciones repetidas pero 150 registros indicaban "N/A" en todas las variables y 100 contenían datos atípicos, por lo que se eliminaron al no representar información que aporte al modelo, la base final contiene 1.913 observaciones.

A continuación se detallan las variables y su incidencia para la clasificación:

Cuadro 4: Clasificación de variables del modelo según uso en los modelos.

Clasificación	Variables	¿Se usó en la clasificación?
Información general de los estudiantes	Número de cédula	No
	Nombre	No
	Discapacidad	Si
	Documentos completos	No
	Estado sistema de beca socioeconómica	No
	Categoría asignada final	Si
	Miembros grupo familiar	No
	Ingreso del estudiante	Si
	RUC empresa estudiante	No
	Empresa estudiante	Si
	RUC estudiante	No
	Auto estudiante	Si
	Avalúo auto estudiante	Si
	Código provincia estudiante	Si
	Código cantón estudiante	No
	Fecha nacimiento estudiante	No
	Nacionalidad estudiante	No
Otros ingresos del estudiante	Si	
Información general del primer representante	Parentesco representante	No
	CI representante	No
	Nombre representante	No
	Código de provincia	No
	Código de cantón	No
	Instrucción representante	No
	Ocupación representante	No
Ingreso representante	Si	

Clasificación	Variables	¿Se usó en la clasificación?
	RUC representante	Si
	Volumen ventas representante	No
	Otros ingresos representante	Si
Información de gastos del estudiante	Gasto arriendo/hipoteca	Si
	Gasto servicios básicos	Si
	Gasto salud	Si
	Gasto alimentación	Si
	Gasto educación	Si
Condiciones especiales	Tipo de vivienda	No
	Condiciones de agua	No
	Fuente de alumbrado	No
	Rango endeudamiento	Si
	Tipo de colegio estudiante	No
	Estado civil estudiante	No
	Internet/tv pagada	Si
	Enfermedad catastrófica del estudiante	Si
	Estudiante con hijos	Si
	Número de hijos	No
Información general del segundo representante.	CI segundo representante	No
	Nombre segundo representante	No
	Nivel de instrucción	No
	Ingreso segundo representante	Si
	RUC segundo representante	Si

Fuente: Dirección de Beneficios Económicos.

Elaborado por: Dayana S.

Como el objetivo no es replicar el modelo que utiliza el Sistema de Beca Diferenciada, se consideró utilizar las variables de “ingreso representante” e “ingreso segundo representante” por separado para ver como influenciaría en la aplicación del algoritmo. Además, se busca realizar una discusión de los resultados en donde se incluya a la variable “otros ingresos representante” que en la base de datos analizada no forma parte del ingreso total. A diferencia del modelo de la PUCE que trabaja con el ingreso del hogar per cápita, el modelo alternativo permitirá ver el comportamiento individual de todas las variables de ingreso.

En la misma línea, la variable “miembros grupo familiar” contenía registros vacíos o con descripciones como "NO REGISTRA", "NO APLICA" o "#N/A", lo cual suponía un impedimento para el cálculo del ingreso per cápita. Considerar la limitación que representa en el modelo.

2.2.2. Transformación y selección de variables

El algoritmo k-medias solo trabaja con valores cuantitativos, pero ya que la investigación se centra en una clasificación socioeconómica es necesario incluir variables que describan este

aspecto social, las mismas son de tipos: binarias y categóricas. Por lo mismo, las variables fueron transformadas a tipo dummy con la fórmula de condición SI de Excel. Véase el cuadro 5 para el detalle de las mismas.

Cuadro 5: Operacionalización de variables.

Dimensión	Variable	Indicador	Tipo	Periodicidad
Educación	Categoría asignada	Documentos de análisis para de asignación categoría	Nominal	Hasta periodo de recategorización
Económica	Ingreso del estudiante	Ingreso del estudiante	Numérica continua	Mensual actualizado
	Avalúo auto estudiante	Bien registrado (automóvil) perteneciente al estudiante.	Numérica continua	Quinquenal actualizado
	Ingreso representante	Ingreso del representante	Numérica continua	Mensual actualizado
	Ingreso segundo representante	Ingreso del segundo representante	Numérica continua	Mensual actualizado
	Otros ingresos representante	Ingresos por otros conceptos aparte del ingreso central del representante.	Numérica continua	Mensual actualizado
	Rango endeudamiento	Si los representantes o el estudiante tienen endeudamiento	Binaria dicotómica (tiene endeudamiento = 1, no tiene endeudamiento = 0).	Mensual actualizado
Financiera	Gasto arriendo/hipoteca	Tabla de amortización de crédito hipotecario de los representantes del estudiante.	Numérica continua	Mensual actualizado
	Gasto servicios básicos	Registros de pagos de servicios básicos	Numérica continua	Mensual actualizado
	Gasto salud	Registros de pagos o valor aproximado del gasto	Numérica continua	Mensual actualizado
	Gasto alimentación	Registros de pagos o valor aproximado del gasto	Numérica continua	Mensual actualizado

Dimensión	Variable	Indicador	Tipo	Periodicidad
	Gasto educación	Registros de pagos de pensión de colegio.	Numérica continua	Mensual actualizado
	Internet/tv pagada	Si los representantes o el estudiante consumen esta clase de servicios	Binaria dicotómica (tiene internet = 1, no tiene internet = 0)	No aplica
Social	Discapacidad (dummy)	Si el estudiante presenta o no discapacidad	Binaria dicotómica (tiene discapacidad = 1, no tiene discapacidad = 0)	Semestral actualizado
	Provincia del estudiante (dummy)	Si el estudiante viene o no de provincia	Binaria dicotómica (viene de provincia = 1, no viene de provincia = 0)	No aplica
	Enfermedad catastrófica del estudiante	Enfermedades catastróficas del estudiante, (en caso de tenerlo).	Binaria dicotómica (tiene enfermedad catastrófica = 1, no tiene enfermedad catastrófica = 0)	Semestral actualizado
	Estudiante con hijos	Si el estudiante tiene o no hijos	Binaria dicotómica (tiene hijos = 1, no tiene hijos = 0)	No aplica
	RUC representante (dummy)	RUC o RISE de los representantes del estudiante.	Binaria dicotómica (registra número de RUC = 1, no registra RUC = 0)	Único actualizado
Laboral	Empresa del estudiante (dummy)	Si el estudiante tiene o no una empresa	Binaria dicotómica (indica tener empresa = 1, indica no	Único actualizado

Dimensión	Variable	Indicador	Tipo	Periodicidad
			tener empresa = 0)	
	RUC segundo representante	RUC o RISE de los representantes del estudiante.	Binaria dicotómica (registra número de RUC = 1, no registra RUC = 0)	Único actualizado

Fuente: Dirección de Beneficios Económicos.

Elaborado por: Dayana S.

La variable binaria dicotómica rango de endeudamiento se definió utilizando la información de la base de datos en donde indica si presenta o no endeudamiento, se busca ver la condición de endeudamiento y es posible que exista sesgo. Se dio un tratamiento similar a las variables RUC representante y RUC segundo representante, en este caso, para 1 se asignaron aquellos registros en donde existe un número de RUC y 0 para lo opuesto.

Para estos casos, se acepta que existen limitaciones en el modelo al no transformar estas variables en base a rangos como el volumen de ventas o la cantidad de endeudamiento. Con respecto a esto último, implica el no saber el tipo de deuda y por ende el nivel de gasto que representa para la familia. Lo ideal sería incluir esta variable con los montos de deuda de tarjetas de crédito, pago de préstamos, etc.; sin embargo, debido al formato en el que se encuentra la información, la misma no pudo ser procesada.

2.3. Aplicación del análisis Clúster con data 2018.

Con el programa R, se usaron los siguientes datos:

- De una muestra de 2.163 observaciones:
 - Se eliminaron 150 casos que no contenían información en la mayoría de las variables.
 - Se eliminaron 100 casos con información atípica.
- Se trabaja con 1.913 observaciones, y 20 variables: 10 variables cuantitativas, 9 variables cualitativas (dicotómicas) y la variable de categoría asignada.

Cuadro 6: Variables del modelo, tipo y denominación en R.

Grupo de variable	Variable	Variable en R
Variable por predecir	Categoría asignada	cat_final
Cuantitativas	Ingreso del estudiante	ingreso_est
	Avalúo auto estudiante	avaluoauto_est
	Ingreso representante	ingreso_rep1
	Otros ingresos representante	otrosingr_rep1

Grupo de variable	Variable	Variable en R
	Gasto arriendo/hipoteca	gastohip_arr
	Gasto servicios básicos	gasto_sbasicos
	Gasto salud	gasto_salud
	Gasto alimentación	gasto_alim
	Gasto educación	gasto_edu
	Ingreso segundo representante	ingreso_rep2
Cualitativas	Discapacidad (dummy)	dDisc
	Empresa del estudiante (dummy)	dEmpresa_est
	Provincia del estudiante (dummy)	dProv_est
	RUC representante (dummy)	dRUC_rep1
	Rango endeudamiento	dEndeudamiento
	Internet/tv pagada	dInternet_Tvpag
	Enfermedad catastrófica del estudiante	dEnfCatr
	Estudiante con hijos	dEst_hijos
	RUC segundo representante	dRUC_rep2

Fuente: Dirección de Beneficios Económicos.

Elaborado por: Dayana S.

Se procede a revisar si existen NAs o casos sin información, el programa indica que no. Posteriormente, se crea un conjunto de datos que no contenga la variable a predecir “categoría asignada”. Una vez que el algoritmo se aplique y se clasifique por clústeres se incluirá de nuevo la variable “categoría asignada” para la comparación de resultados entre el modelo PUCE y el modelo alternativo.

Normalmente, para un conjunto de datos en el que no hay un conocimiento previo sobre el número de grupos en los que se puede clasificar la información, se usa el estadístico F que proporciona el número óptimo de grupos. Debido a que se sabe con anterioridad que la PUCE tiene cinco clasificaciones no es necesario este cálculo, sin embargo se presentan los resultados obtenidos:

```
> for (i in 1:10) {
+   SCDG[i] <- sum(kmeans(ndata,i,nstart = 100)$withinss)
+ }
> (SCDG[2]-SCDG[3])/(SCDG[3]/n-2-1)
[1] 378.6857
> (SCDG[3]-SCDG[4])/(SCDG[4]/n-3-1)
[1] 231.42
> (SCDG[4]-SCDG[5])/(SCDG[5]/n-4-1)
[1] 240.8192
> (SCDG[5]-SCDG[6])/(SCDG[6]/n-5-1)
[1] 215.3439
> (SCDG[6]-SCDG[7])/(SCDG[7]/n-6-1)
[1] 224.3401
```

La prueba F, según (Peña, 2002) determina lo siguiente:

$$F = \frac{SCDG(G) - SCDG(G + 1)}{SCDG(G + 1) / (n - G - 1)}$$

siendo *SCDG* la suma de cuadrados, *G* el número de grupos y *n* el número de observaciones, se recomienda agregar un grupo adicional si el resultado es igual o mayor a 10. Los resultados de la prueba F para el conjunto de datos sugerían clasificar en más de diez grupos, dado que se busca comparar con el modelo de la PUCE, se decide correr el algoritmo con cinco grupos (Arroyo Ruales, 2017):

```
> set.seed(123)
> kmeans.est <- kmeans(ndata,centers = 5, nstart = 120)
> kmeans.est
K-means clustering with 5 clusters of sizes 281, 1084, 93, 393, 62
```

La clasificación arroja una gran concentración de datos en un grupo, el 57%. Mientras que para los grupos restantes, la cantidad de observaciones asignadas está relativamente equilibrada. Es probable que si se utilizaba el número de grupos resultante del cálculo de la prueba F, hubiese existido una dispersión del grupo de 1.084 observaciones en más clústeres. Asimismo, por las características de los datos, es también posible que el mayor número de observaciones perteneciese al mismo clúster.

Cada clúster contiene el siguiente número de observaciones:

Cuadro 7: Número de estudiantes por categoría según modelo cuantitativo/cualitativo.

N.º de clúster en R	Categoría	nº
2	A	1.084
4	B	393
1	C	281
5	D	62
3	E	93

Fuente y elaboración: Dayana S.

El modelo tuvo cinco iteraciones para llegar a la clasificación final, es decir, cada observación está a la distancia euclídea óptima de cada centroide.

```
> #Número de iteraciones
> kmeans.est$iter
[1] 5
```

Medida de calidad del modelo

Between_SS/total_SS es un determinante de calidad del modelo, significa la intra separación del clúster y la inter separación entre clústeres, en este caso, es del 53,9%. Como indica la teoría del método, tanto este estadístico como el ratio de desemejanza buscan maximizar su

resultado, pues mientras más cerca esté del 100% mejor será la calidad del modelo (Attewell, Monaghan, & Kwong, 2015).

Este estadístico puede incrementarse si se aumenta el número de grupos. Dado que, en el siguiente capítulo el objetivo es comparar los resultados con las cinco categorías ya asignadas por la PUCE, no se aumenta el número de clústeres, que podría o no mejorar este estadístico.

```
within cluster sum of squares by cluster:
[1] 435638643 683859706 372172662 515647927 200080373
(between_ss / total_ss = 53.9 %)
```

Cuadro 8: Medias por variable de cada clúster modelo cuantitativo.

Cluster means:						
	dDisc	ingreso_est	dEmpresa_est	avaluoauto_est	dProv_est	
1	0.01423488	16.90374	0.01423488	3.217082	0.1601423	
2	0.01660517	27.81337	0.01660517	23.756614	0.1697417	
3	0.01075269	15.55914	0.01075269	0.000000	0.2580645	
4	0.01526718	15.94595	0.01526718	28.569975	0.2264631	
5	0.00000000	22.16952	0.00000000	0.000000	0.3064516	
	ingreso_rep1	dRUC_rep1	otrosingr_rep1	gastohip_arr		
1	1125.6651	0.3451957	414.8153	271.7855		
2	711.1827	0.3809963	230.2139	139.9496		
3	2847.0357	0.3118280	534.3754	550.7519		
4	1933.2406	0.5343511	183.4517	371.6737		
5	4443.5427	0.6129032	294.0161	705.4816		
	gasto_sbasicos	gasto_salud	gasto_alim	gasto_edu	dEndeudamiento	
1	100.67110	164.44313	368.7491	461.9288	0.9430605	
2	79.46478	86.78979	259.1208	232.1958	0.8570111	
3	135.00548	287.11989	634.2075	705.5719	0.9892473	
4	125.52982	196.42445	462.0053	542.7497	0.9567430	
5	173.46000	381.13968	718.7792	743.0192	0.9516129	
	dInternet_Tvpag	dEnfCatr	dEst_hijos	ingreso_rep2	dRUC_rep2	
1	0.9395018	1	0.014234875	1479.9414	0.6654804	
2	0.8404059	1	0.029520295	206.6663	0.5009225	
3	0.9784946	1	0.010752688	3179.9808	0.7526882	
4	0.9541985	1	0.005089059	240.9273	0.6132316	
5	1.0000000	1	0.032258065	422.6392	0.7580645	

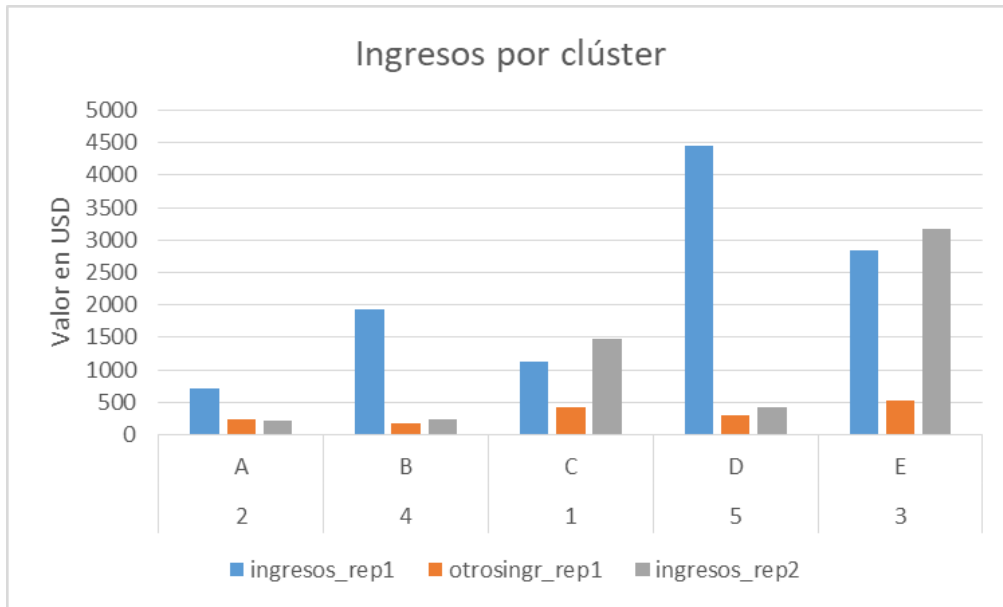
Fuente: Dirección de Beneficios Económicos.

Elaborado por: Dayana S.

Los resultados del cuadro 9 indican que, por ejemplo, la media de ingreso del estudiante del clúster 2 es de \$27,81. Una interpretación para las variables cualitativas no se puede realizar pues contienen valores binarios en donde no tiene sentido el cálculo de una media, siendo esta una de las razones por las que el algoritmo k-medias solo trabaja con valores.

Para la discusión de los resultados del cuadro 9 se analiza la variable de ingresos:

Gráfico 11: Distribución de ingresos por clúster del modelo cuantitativo/ cualitativo.



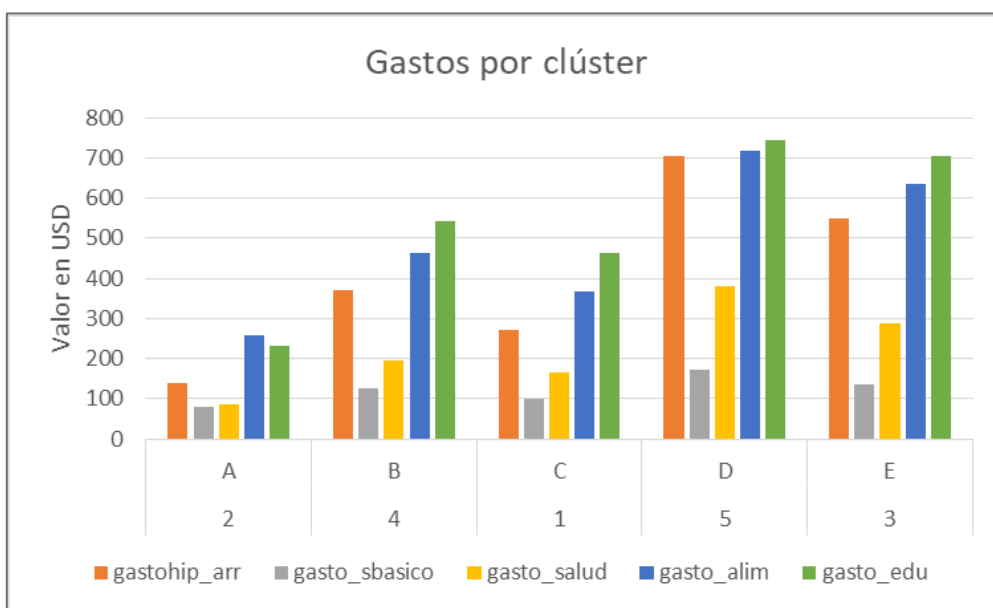
Fuente: Dirección de Beneficios Económicos.

Elaborado por: Dayana S.

Para el clúster 2, los ingresos de los representantes del estudiante son los más bajos y no superan los \$1.000,00, por lo que se ubica en la categoría A. En el caso del clúster 3, los ingresos por separado no superan los \$3.500,00, sin embargo, su suma, \$6.561,39, sobrepasa a los ingresos totales del clúster 5, \$5.160,19, y se asigna a la categoría E.

Se asigna el clúster 5 a la categoría D, a pesar de que visualmente, este clúster presenta el ingreso del representante 1 más alto, es interesante observar cómo se comporta la variable del ingreso del representante 2 y los otros ingresos del representante 1, que comparado con la categoría E son notablemente menores.

Gráfico 12: Distribución de gastos por clúster del modelo cuantitativo/ cualitativo.



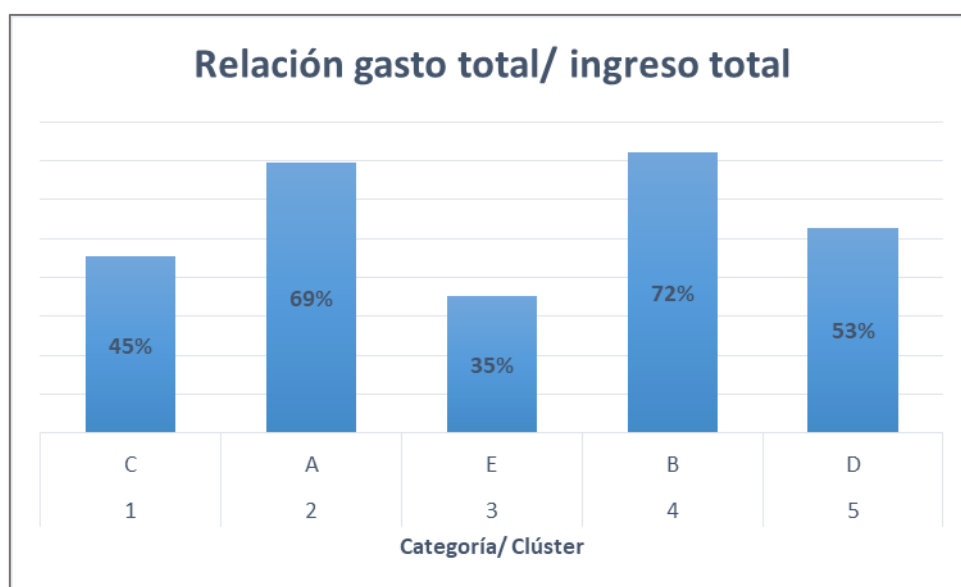
Fuente: Dirección de Beneficios Económicos.

Elaborado por: Dayana S.

Al igual que en el gráfico 11, los gastos del clúster 2 son los más bajos y no sobrepasan los \$300,00, por lo cual se ubica en la categoría A. En el clúster 5, todos los tipos de gasto sobrepasan a los de los clústeres restantes, y se lo asigna a la categoría D porque es una disminución mayor en los ingresos de la familia comparada con los gastos del clúster 3. Entonces, en la categoría E se encuentran los hogares que tienen un mayor ingreso y menor gasto con respecto a la categoría D que tiene un ingreso elevado pero con menos gastos.

Con relación a la distribución del tipo de gasto, visualmente el mayor gasto de las familias es en educación con un rango de entre \$450,00 y \$730,00. El segundo componente más importante alimentación, con un gasto de entre \$250,00 y \$710,00. Seguido por el gasto en hipoteca/arriendo que va de entre \$150,00 y \$700,00. Mientras que gastan menos en servicios básicos y salud.

Gráfico 13: Porcentaje de gasto total sobre el ingreso total del modelo cuantitativo/ cualitativo.



Fuente: Dirección de Beneficios Económicos.

Elaborado por: Dayana S.

El porcentaje de gasto con respecto del ingreso de la categoría B es del 72%, lo que indica que los gastos en salud, educación, arriendo/ hipoteca, alimentación y servicios básicos representan la mayor parte de la estructura financiera de la familia con respecto a sus ingresos y básicamente, gastan el 72% del ingreso total, \$2.357,61. En el clúster 3, el 35% de los ingresos se utilizan en cubrir los tipos de gastos mencionados. Este análisis indica que para los hogares de las categorías A, B y D; los gastos consumen la mayor parte de sus ingresos. Considerar la limitación de que esta estructura financiera no incluye gastos por concepto de pago de deuda.

2.3.1. Aplicación del algoritmo k-medias con variables cuantitativas

Ya que la aplicación del k-medias es específicamente para valores y variables cuantitativas, se volvió a correr el modelo solo con las variables cuantitativas y con las consideraciones anteriores.

```
> set.seed(123)
> kmeans.estc <- kmeans(ndata,centers = 5, nstart = 120)
> kmeans.estc
K-means clustering with 5 clusters of sizes 1087, 93, 390, 62, 281
```

Resultando que cada clúster, con este modelo tiene el siguiente número de elementos:

Cuadro 9: Número de estudiantes por categoría según modelo cuantitativo.

N.º de clúster en R	Categoría	nº
1	A	1.087
3	B	390
5	C	281
4	D	62
2	E	93

Fuente y elaboración: Dayana S.

La clasificación óptima llegó después de cuatro iteraciones:

```
> #Número de iteraciones
> kmeans.estc$iter
[1] 4
```

La medida de calidad ($\text{between_SS}/\text{total_SS}$) del modelo se mantiene como en el modelo anterior pues no se incrementó el número de grupos:

```
within cluster sum of squares by cluster:
[1] 685947961 372172601 513559422 200080329 435638437
(between_SS / total_SS = 53.9 %)
```

En los resultados de las medias de cada clúster, esta vez las medias de las variables son interpretables:

Cuadro 10: Medias por variable de cada clúster modelo cuantitativo.

Cluster means:					
	ingreso_est	avaluoauto_est	ingreso_rep1	otrosingr_rep1	
1	28.09171	23.691049	713.0451	229.5786	
2	15.55914	0.000000	2847.0357	534.3754	
3	15.07887	28.789744	1937.4504	184.8628	
4	22.16952	0.000000	4443.5427	294.0161	
5	16.90374	3.217082	1125.6651	414.8153	
	gastohip_arr	gasto_sbasicos	gasto_salud	gasto_alim	gasto_edu
1	140.5606	79.43771	86.86519	259.1569	231.9388
2	550.7519	135.00548	287.11989	634.2075	705.5719
3	371.7531	125.95962	197.05764	463.4652	545.8549
4	705.4816	173.46000	381.13968	718.7792	743.0192
5	271.7855	100.67110	164.44313	368.7491	461.9288
	ingreso_rep2				
1	206.0959				
2	3179.9808				
3	242.7806				
4	422.6392				
5	1479.9414				

Fuente: Dirección de Beneficios Económicos.

Elaborado por: Dayana S.

Con lo anterior podemos inferir lo siguiente de las variables más relevantes:

- La media más alta del ingreso del representante 1 del estudiante de \$4.443,54 dólares es la correspondiente a la categoría D, mientras que la más baja de \$713,04 dólares es de la categoría A.
- La media más alta del ingreso del representante 2 del estudiante de \$3.179,98 dólares es la correspondiente a la categoría E, mientras que la más baja de \$206,09 dólares es de la categoría A.
- La media más alta del gasto en hipoteca/ arriendo de \$705,48 dólares es la correspondiente a la categoría D, mientras que la más baja de \$140,56 dólares es de la categoría A.
- La media más alta del gasto en servicios básicos de \$173,46 dólares es la correspondiente a la categoría D, mientras que la más baja de \$79,43 dólares es de la categoría A.

La técnica de K-medias fue aplicada a dos conjuntos de datos. En el primero con variables cuantitativas y cualitativas, que coincide con el objetivo de clasificación socioeconómica. Sin embargo, por las limitaciones del algoritmo por el uso exclusivo de variables cuantitativas, se trató con un modelo con las características referidas.

Para los dos conjuntos de variables se muestra una clasificación similar de estudiantes por categoría y tienen estadísticos de calidad de modelo idéntico. Esto último, se entiende ya que se vería una mejora debido al incremento del número de grupos, más no porque se deba modificar la información. El ejercicio ha servido para hacer una comparación de los dos conjuntos de datos.

Al final, lo descrito ha permitido alcanzar el objetivo de probar una nueva técnica de clasificación, observar sus resultados y sugerir el uso de esta técnica con las mejoras y recomendaciones correspondientes. Esto último dado las limitaciones en el procesamiento de datos y en la transformación de variables aplicadas en el modelo.

Además, se demuestra que trabajar con diferentes grupos de variables no influye en como clasifica el algoritmo, pues el segundo modelo, a pesar de haber trabajado solo con variables cuantitativas, arrojó resultados parecidos a los del modelo cuantitativo/ cualitativo. Se demuestra que incluir grupos de variables dicotómicas no influye en gran manera.

Capítulo III: Incidencia de los resultados del modelo alternativo

3.1. Contraste entre los resultados de categorización del modelo alternativo con resultados del Sistema de Beca Socioeconómica 2018

A continuación se comparan los modelos alternativos detallados en el apartado anterior, el modelo cuantitativo/cualitativo y el modelo cualitativo. Se presentan los resultados y también se hace una comparación con los resultados del modelo de categorización de la PUCE. En resumen los resultados de los dos modelos alternativos son los siguientes:

Cuadro 11: Comparación de estadísticos y características de los modelos alternativos.

Modelo	Iteraciones	between_SS /total_SS	Medias clústeres	N° variables usadas
Cualitativo/ cuantitativo	5	53,9%	No se puede interpretar con variables cualitativas	20
Cuantitativo	4	53,9%	Se puede interpretar	10

Fuente y elaboración: Dayana S.

A pesar de la limitación del modelo cuantitativo, de no poder analizar las variables cualitativas del conjunto de datos, la medida de calidad "between_SS/total_SS" no varía. La medida tendría un cambio si se modificara el número de grupos de clasificación, pero dado que se busca hacer una comparación con el modelo de categorización de la PUCE-Q, se mantiene los cinco grupos o categorías establecidos.

Cuadro 12: Resultados por número de estudiantes en cada categoría.

Categoría	Modelo PUCE	Modelo Cualitativo/Cuantitativo	Modelo Cuantitativo
A	461	1.084	1.087
B	509	393	390
C	558	281	281
D	211	62	62
E	174	93	93

Fuente: Dirección de Beneficios Económicos.

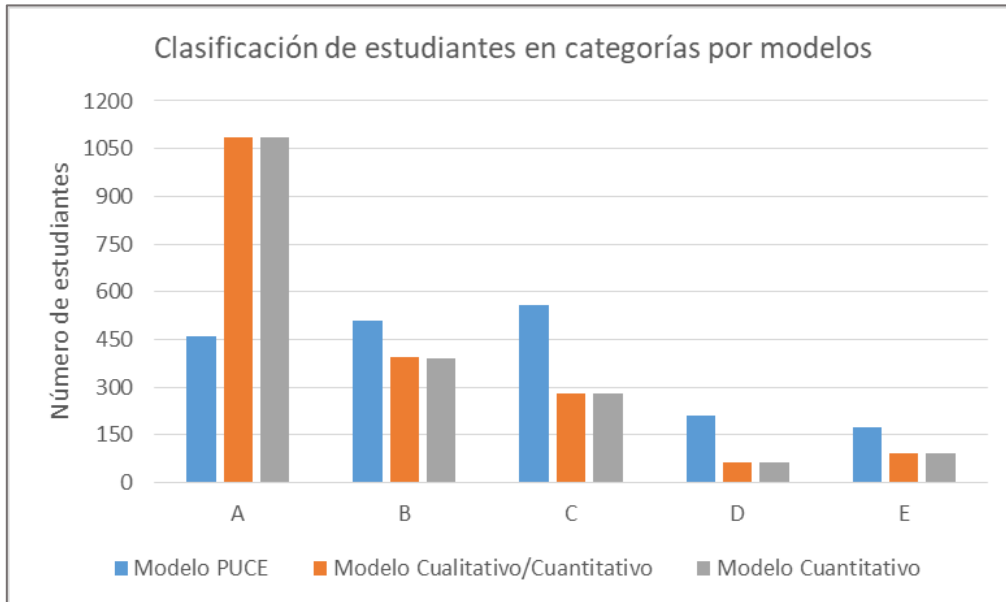
Elaborado por: Dayana S.

En el cuadro 11 los resultados entre los modelos alternativos son los mismos con respecto a la clasificación en las categorías C, D y E. Con A y B, se nota una diferencia mínima de 3 registros que se movieron entre los grupos mencionados. Al probar con los dos conjuntos de variables, y obtener los mismos resultados, se concluye que las características de los datos y

de cada observación tienen un peso importante en la iteración del algoritmo, más que si se añadan, se quiten o se jueguen con distintos grupos de variables.

De manera visual, se observa la comparación de la clasificación de los 3 modelos:

Gráfico 14: Clasificación de estudiantes en las cinco categorías por modelo.



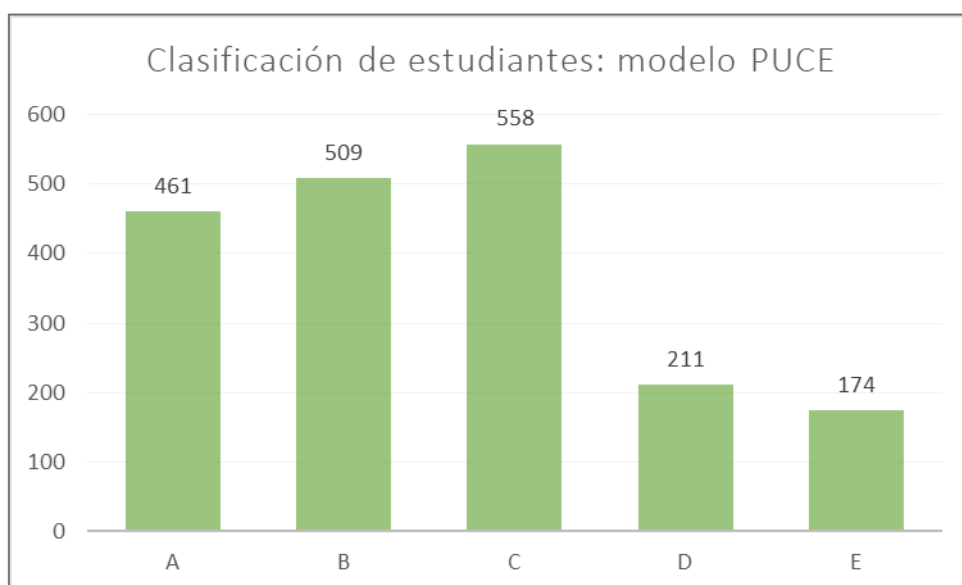
Fuente: Dirección de Beneficios Económicos.

Elaborado por: Dayana S.

Los modelos alternativos clasifican a más estudiantes en la categoría A, mientras que el modelo de la PUCE asigna en la categoría C. Por el contrario, los modelos alternativos categorizan a menos observaciones en la categoría D y el modelo de la PUCE lo hace en la categoría E, coincidiendo en realizar esta asignación en las categorías con valores elevados de aranceles.

En un análisis más detallado de la clasificación de cada modelo:

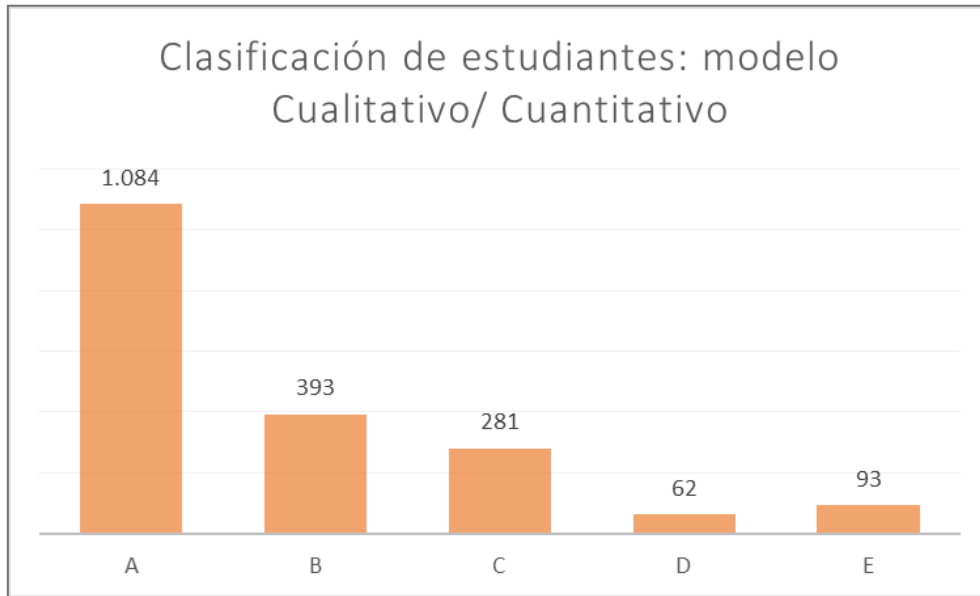
Gráfico 15: Clasificación de estudiantes por categoría: modelo PUCE.



Fuente: Dirección de Beneficios Económicos.
Elaborado por: Dayana S.

Se evidencia que el 65% de la clasificación de los estudiantes se concentra en las tres primeras categorías A, B y C. Siendo la categoría C la que más estudiantes engloba, y F la que menos.

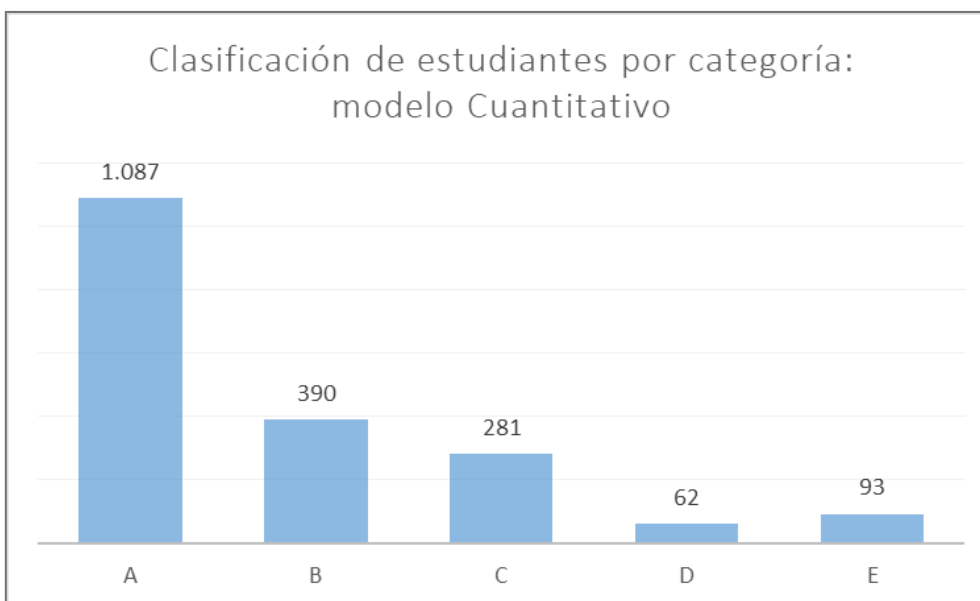
Gráfico 16: Clasificación de estudiantes por categoría: modelo Cualitativo/ Cuantitativo.



Fuente: Dirección de Beneficios Económicos.
Elaborado por: Dayana S.

En este caso, la clasificación de este modelo presenta una tendencia descendente, pues la categoría A contiene el 57% de los elementos. Por otra parte, la categoría D tiene la menor proporción de estudiantes.

Gráfico 17: Clasificación de estudiantes por categoría: modelo Cuantitativo.



Fuente: Dirección de Beneficios Económicos.

Elaborado por: Dayana S.

Al igual que la anterior clasificación, el 57% de la distribución de los datos se agrupa en la categoría A y la menor en la categoría D.

Como la clasificación con los modelos de k-means son similares, para la discusión de resultados se los analiza en conjunto y se denomina como modelos alternativos. Para una óptica distinta se agrupó a las categorías en tres grupos, Con el objetivo de mostrar esto, se presenta las proporciones de estudiantes en cada grupo y por modelo, véase el cuadro 12:

Cuadro 13: Resultados por proporción de agrupación de categorías.

Modelo	Estudiantes en categorías A y B	Estudiantes en categorías C y D	Estudiantes en categoría E
PUCE	51%	40%	9%
Modelos alternativos	77%	18%	5%

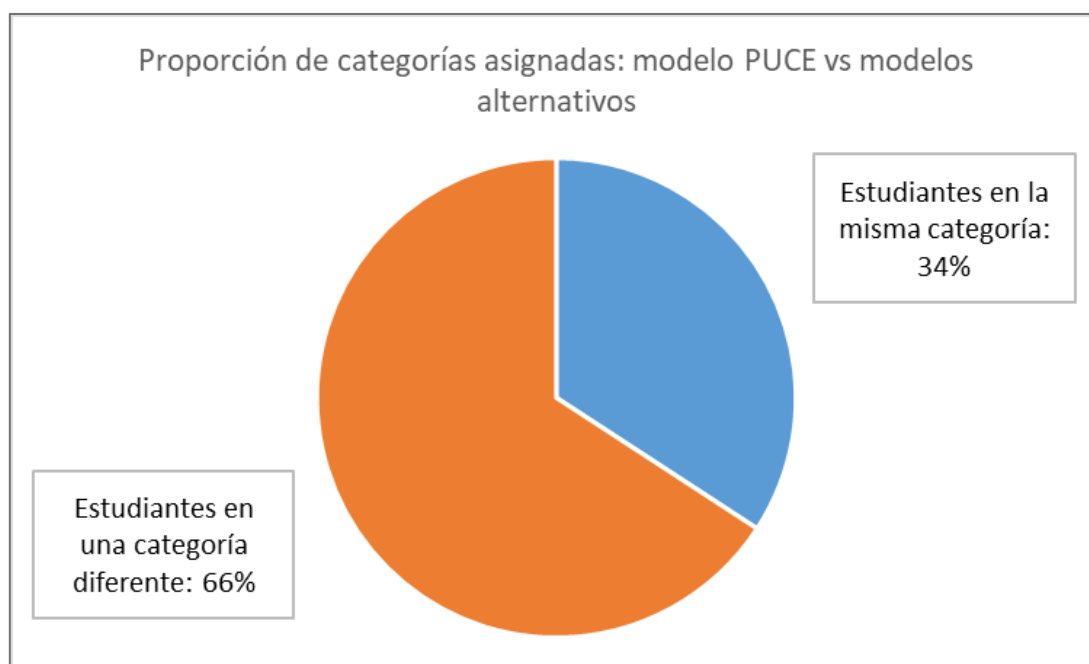
Fuente: Dirección de Beneficios Económicos.

Elaborado por: Dayana S.

En lo visualizado anteriormente, se evidencia que podría haber una tendencia a clasificar a los estudiantes dentro de las primeras cuatro categorías. Lo anterior indica que el modelo PUCE tiende a ubicar a un poco más de la mitad de sus estudiantes en las categorías A y B. Asimismo, los modelos alternativos coinciden en ubicar al 77% de las observaciones en el rango de las mismas categorías.

En otro orden de ideas y con el fin de comparar la asignación de categorías en cada modelo, se realiza el siguiente análisis:

Gráfico 18: Diferencias en categorías asignadas: modelo PUCE vs modelos alternativos.



Fuente: Dirección de Beneficios Económicos.

Elaborado por: Dayana S.

Dado que los resultados de clasificación de los modelos cuantitativo/ cualitativo y cuantitativo tienen resultados similares la proporción es la misma por lo que se hace una comparación de los modelos analizados con el de la PUCE y existe una diferencia en la asignación de categorías del 66% y una similitud del 34%.

La información de las diferencias de categorización entre modelos de cada observación no se incluye por su longitud. El detalle de la misma estaría disponible en caso de ser solicitada.

3.2. Consideración de otros factores que afectan a la asignación de recursos con el modelo de categorización

Ciertas consideraciones se toman en cuenta durante la disertación acerca los criterios de clasificación socioeconómica. Sin embargo, hay varios factores que son importantes en la asignación de recursos, los cuales no están considerados en el modelo o sus fundamentos. Esto debido a la dificultad para calcular estos factores e incorporarlos en los modelos alternativos, así como de especificar a qué tipo de eficiencia se iba a referir el análisis.

Eficiencia en manejo de recursos universitarios

El manejo de los recursos viene atado a hacerlo de manera eficiente, minimizando las desigualdades que se pueden presentar cuando los mismos se asignan. Básicamente, como indica el óptimo de Pareto, mejorando el bienestar de una persona sin perjudicar el de otra. Los recursos mencionados en el caso de la unidad de análisis son principalmente, monetarios y proporcionados por agentes privados. Esto brinda una perspectiva diferente sobre su uso hacia lo que se denomina productos universitarios, como infraestructura y lo relacionado a educación; estudiantes graduados, investigaciones realizadas, etc. (García de Fanelli, 2005)

La consecución de los productos universitarios de manera eficiente, implica un excedente de recursos que invita a su ahorro o inversión que ayuden a mejorar el sistema universitario. La eficiencia es un eje importante para que; un sistema universitario, organización, empresa; cumpla sus objetivos relacionados con los recursos y limitaciones que presenten (García de Fanelli, 2005).

Sin embargo, es difícil aplicar estas concepciones teóricas de eficiencia en modelos numéricos. Es por eso que esta clase de factores podría influir de cierta manera el modelo o el proceso del mismo, a pesar de que es complicado plasmarlo en la metodología desarrollada.

Aproximación de costo económico por modelo

Otro elemento a considerar es el costo que le representa a la universidad el aplicar un sistema de beneficios económicos como el SBS. De igual manera, se puede establecer una comparación con el costo de cada modelo alternativo y lo que implica en materia económica el ubicar a un estudiante en una categoría. A fin de demostrar lo anterior, se realiza un cálculo aproximado del valor económico de clasificar con los resultados de los tres modelos: PUCE-Q, cuantitativo/ cualitativo y cuantitativo. Para esto, se realizó lo siguiente:

- Empatar observación por número de cédula con el código de carrera al que pertenece.
- Agrupar la información por carrera en las seis clasificaciones por área de la tabla de aranceles de Dirección de Beneficios Económicos⁴.
- Extrapolar el número de estudiantes de la muestra (1.913 observaciones) con el número de estudiantes total o población de la PUCE-Q (10.496 observaciones):
 - Se identifica el código de carrera de cada observación de la muestra para definir la carrera que el estudiante está cursando.
 - Se agrupa las observaciones con su carrera identificada por categoría.
 - Se determina la proporción de estudiantes por carrera con respecto al total de la muestra.
 - Para extrapolar, esta proporción se multiplica por la población de estudiantes de la carrera clasificados por categoría.
 - Se clasifica la información por carrera en las seis áreas de conocimiento de la tabla de aranceles de DBE.
 - La cantidad extrapolada de estudiantes clasificados se multiplica, según su categoría, por el arancel tipo 1 de cada área de conocimiento.
 - Se suma el valor monetario de cada categoría para obtener el coste económico de los modelos alternativos.

Resultando lo siguiente:

Cuadro 14: Costo económico aproximado de los modelos.

Modelo PUCE-Q	Modelos alternativos
\$14.044.230,00	\$15.083.407,00

Fuente: Dirección de Beneficios Económicos.

Elaborado por: Dayana S.

El costo de implementar una clasificación socioeconómica mediante el SBS le representa a la PUCE-Q un total de \$14.044.230,00. Es decir, esta cantidad deja de percibir la universidad al realizar el descuento por categoría a todos los estudiantes. Difiere con el costo de los modelos alternativos, \$15.083.407,00; lo cual es entendible puesto que los modelos alternativos clasificaron a los estudiantes en un 57% en la categoría A.

Por la clasificación que muestran, agrupando las observaciones en las dos primeras categorías, los modelos alternativos no significarían un aporte en términos de sostenibilidad financiera. Los ingresos que se obtendrían al ubicar al 77% de los estudiantes en las categorías A y B disminuirían, por el descuento de hasta el 55% y 45% que presentan en el pago semestral, respectivamente. Por ende afectando a la estructura presupuestal de la PUCE en uno de sus rubros más importantes.

Como conclusión, en vista de la utilización de diferentes técnicas es interesante evidenciar las similitudes en los resultados de los modelos tratados. Resulta claro que la distribución de los datos iba a ser diferente, al igual que la asignación de categorías. Por otra parte, se indica a la medida de eficiencia universitaria a los factores que influyen en la asignación de recursos que no necesariamente se ve de manera explícita en los modelos, pero tampoco deja de ser una consideración influyente dentro de los mismos. Se incluye además un cálculo del costo económico aproximado aplicado a cada modelo y las posibles implicaciones en la estructura financiera de la PUCE-Q.

⁴ Ver detalles de tabla de aranceles en el Anexo n.º 5.

Conclusiones

El trabajo de investigación ha cumplido con el objetivo de realizar una propuesta alternativa de categorización socioeconómica de los estudiantes con una técnica de clasificación perteneciente al área de minería de datos. Esto permitió conocer sobre la aplicación del algoritmo k-medias en el área educativa. Sobre todo, se rescata además el conocimiento adquirido sobre el modelo de categorización de la PUCE.

En la misma línea, los modelos desarrollados se asemejan en la utilización de dimensiones parecidas que comprenden la información del estudiante y su entorno. En cuestión, esto se refiere a dimensiones de ingresos, gastos y aquellas que engloban condiciones de vulnerabilidad. Lo último indicando que la mayoría de clasificaciones socioeconómicas persiguen lo mismo y con recursos similares, aunque con técnicas distintas.

El modelo de Sistema de Beca Socioeconómica se desarrolla en base a características similares. Procura un análisis integral de las condiciones económicas y sociales del estudiante y sus representantes. Por lo mismo, y debido a su estructura, se enfoca en la dimensión de vulnerabilidad como principal reflejo de la realidad del entorno del estudiante. Comprendiendo esto, se entiende que la mayor concentración de casos se encuentra en las categorías A, B y C; las mismas que se refieren a aquellas con el mayor descuento en el pago por educación en la PUCE. Algo semejante se trató de incluir en los modelos de clasificación resultantes de aplicar el algoritmo k-medias. Se manejaron variables cuantitativas y cualitativas siguiendo la idea del modelo de la PUCE. La creación del modelo exclusivamente cuantitativo viene de la concepción teórica que se tiene sobre el algoritmo k-medias, que recomienda el uso de este tipo de variables.

El modelo cuantitativo/ cualitativo usó 20 variables que contemplan información económica; sobre ingresos y gastos, e información social; como discapacidad y enfermedades catastróficas. Resulta que la clasificación agrupa al 57% de estudiantes de la muestra en la categoría A y solo al 5% en la categoría E. De nuevo, el modelo cuantitativo clasificó con 10 variables cuantitativas que engloban solo los aspectos económicos del análisis y tiene resultados similares al primer modelo mencionado. Se reconoce además que el modelo tiene limitaciones en el manejo y tratamiento de las variables, además de haber trabajado solo con datos auto reportados.

Otra comparación aplicada, el cálculo del costo económico de cada modelo. En resumidas cuentas, el modelo de la PUCE-Q y los modelos alternativos expresan un resultado diferente. Siendo los últimos más caros de implementar, además de demostrar una clasificación sesgada hacia las dos primeras categorías, demostrando que su aplicación podría suponer insostenibilidad financiera en las cuentas universitarias.

Recomendaciones

La recomendación más importante de este o cualquier otra propuesta de modelo alternativo siempre será el actualizarlo e ir probando con diferentes variables que aporte o describan mejor las condiciones socioeconómicas del estudiante. Siempre se puede mejorar con miras a ofrecer al estudiante una mejor condición de acceso y culminación de estudios, al igual que un sistema de apoyo por parte de la universidad que garantice las mejoras necesarias y constantes en sus procesos de categorización y gestiones financieras.

En el capítulo 3, la eficiencia es considerada como componente que afecta a la asignación de recursos. Se recomienda una mejora de estos modelos ajustados a las consideraciones descritas de este elemento faltante, ya que esto no se encuentra de manera explícita en la aplicación de los modelos alternativos, a pesar de que se considera a estos conceptos como parte fundamental del manejo de recursos.

La modelación tendrá variaciones considerables cuando se agregan nuevas variables que vale la pena analizar. En este caso no es distinto, el incluir variables o medirlas de distinta manera puede evidenciar resultados relevantes. La inclusión de estas variables puede venir desde la óptica del investigador, así como por los objetivos de la investigación.

De hecho, en el trabajo de disertación de Denisse Agama de la carrera de gestión social en donde analiza el tema de inclusión social en el SBS, concluye con la adición de dos nuevas variables. Las mismas fueron obtenidas por resultados de encuestas y son “estudiante soltera embarazada” y “estudiante soltero/a – fuera del núcleo familiar” (Agama Galárraga, 2014).

Con respecto a la última agrego mi opinión como estudiante de provincia, considero que los gastos en alimentación, transporte, vivienda (arriendo), y personales; no son recogidos en su totalidad dentro del modelo de SBS cuando si representa un peso importante en la estructura financiera de las familias con estos casos.

A pesar de que se trata de subsanar estas cuestiones con la entrevista realizada por Dirección de Beneficios Económicos, aún no se dimensiona de manera adecuada el esfuerzo que los estudiantes y sus familias realizan. Esta es la realidad del 18% de los estudiantes de provincia del año 2018 en la PUCE-Q⁵.

Otro análisis sobre categorización estudiantil en la Escuela Politécnica Nacional lo realiza Wilson Arroyo, en donde utiliza variables similares a las usadas en los modelos estudiados en esta disertación. Un aporte destacable es la inclusión de la variable de “valor del terreno”. Si bien no es una propiedad como una casa o un edificio, si es una parte importante de un patrimonio familiar (Arroyo Ruales, 2017).

⁵ Cálculo realizado con información de la matriz de Estudiantes de la Dirección General de Estudiantes.

Referencias Bibliográficas

- Abramo, L., Cecchini, S., & Morales, B. (2019). Programas sociales, superación de la pobreza e inclusión laboral: Aprendizajes desde América Latina y el Caribe. *Libros de la CEPAL*(155). Obtenido de https://www.cepal.org/sites/default/files/events/files/c1900117_web.pdf
- Agama Galárraga, D. Y. (2014). Inclusión Económica y Social en la Educación Superior Caso: Pensión Diferenciada de la Unidad Académica de Trabajo Social de la PUCE. Obtenido de <http://repositorio.puce.edu.ec/bitstream/handle/22000/6347/102.17.000819.pdf?sequence=4&isAllowed=y>
- Alkire, S., Roche, J. M., Santos, M. E., & Seth, S. (2011). *Multidimensional Poverty Index 2011*. Oxford Poverty & Human Development Initiative (OPHI). University of Oxford. Obtenido de <http://ophi.org.uk/wp-content/uploads/OPHI-MPI-Brief-2011.pdf>
- Amat, J. (Septiembre de 2017). Clustering y Heatmaps: Aprendizaje no Supervisado. Obtenido de https://www.cienciadedatos.net/documentos/37_clustering_y_heatmaps
- Arroyo Ruales, W. E. (Junio de 2017). *Metodología para la clasificación socioeconómica de los estudiantes de la Escuela Politécnica Nacional y su incidencia en la asignación de becas y cobros de aranceles*.
- Attewell, P., Monaghan, D., & Kwong, D. (2015). *Data Mining for the Social Sciences: An Introduction* (1 ed.). University California Press. Obtenido de <https://www.jstor.org/stable/10.1525/j.ctt13x1gcg>
- Blaug, M. (Octubre de 1993). La Economía en la Educación Superior. *Revista de la Educación Superior*, 22(88), 1-13. Obtenido de http://publicaciones.anuies.mx/pdfs/revista/Revista88_S2A4ES.pdf
- Callan, T., & Keane, C. (2009). Non-Cash Benefits and the Distribution of Economic Welfare. *Institute for the Study Labor (IZA)*. Obtenido de <http://ftp.iza.org/dp3954.pdf>
- Cardona Acevedo, M., Montes Gutiérrez, I. C., Vásquez Maya, J., Villegas Gonzáles, M. N., & Brito Mejía, T. (Abril de 2007). Capital Humano: Una Mirada desde la Educación y la Experiencia Laboral. *Universidad EAFIT: Revistas Académicas*(56), 1-30. Obtenido de <http://publicaciones.eafit.edu.co/index.php/cuadernos-investigacion/article/view/1287>
- Carrasco Delhy, L. F. (Julio de 2014). Construcción de un modelo de clasificación socioeconómica para el descuento de las colegiaturas de los estudiantes de la FLACSO.
- Comisión de becas PUCE-SI. (2012). *Manual de procedimientos para la aplicación del Sistema de Pensión Diferenciada en la PUCE-SI*. Obtenido de http://www.pucesi.edu.ec/phocadownloadpap/manuales/3_ManualSPD2012APR.pdf
- Destinobles, A. G. (2006). *El Capital Humano en las Teorías del Crecimiento Económico*. Chihuahua, México: Eumed.
- Dirección General de Estudiantes. (2018). *Matriz Integral de Estudiantes*.

- García de Fanelli, A. M. (2005). Universidad, organización e incentivos. desafíos de la política de financiamiento frente a la complejidad institucional. Argentina: Miño y Dávila. Recuperado el 30 de Mayo de 2018, de Proquest: <https://ebookcentral.proquest.com>
- Gerrings, J. (2004). *What is a case study and what is good for?* American Political Science Association.
- González, I., & Muñoz, G. (2018). *Diagnóstico y Optimización del Sistema de Beca Socioeconómica (SBS) 2018*.
- Hernández, R., Fernández, C., & Baptista, P. (2010). *Metodología de la investigación*. México: McGraw Hill /Interamericana Editores S.A. de C.V.
- Horna, L., Guachamín, M., & Guerrero, C. (2009). Análisis del mercado del sector "Construcción" bajo un enfoque de concentración económica, en el caso ecuatoriano, durante el periodo 2000-2008. *Revista Politécnica*, 30(1), 217-229.
- INEGI. (2020). *Clasificación de Tipo de Discapacidad - Histórica*. México. Obtenido de https://www.inegi.org.mx/contenidos/clasificadoresycatalogos/doc/clasificacion_de_tipo_de_discapacidad.pdf
- Krupp, S. (Septiembre de 1968). Axioms of Economics and the Claim to Efficiency. *Journal of Economic Issues*, 2(3), 275-282. Recuperado el 25 de abril de 2019, de <https://www.jstor.org/stable/4223956>
- LOES. (2018). *Consejo de Educación Superior*. Recuperado el 30 de Mayo de 2019, de Ley Orgánica de Educación Superior: <http://www.ces.gob.ec/documentos/Normativa/LOES.pdf>
- LSE. (2010). Poverty, inequality and public cash transfers: lessons from Latin America. *LSE Research Online*. Obtenido de [http://eprints.lse.ac.uk/36840/1/Poverty%20inequality%20and%20public%20cash%20transfers%20\(lsero\).pdf](http://eprints.lse.ac.uk/36840/1/Poverty%20inequality%20and%20public%20cash%20transfers%20(lsero).pdf)
- Ministerio del Trabajo. (2018). Ministerio del Trabajo establece Salario Básico Unificado 2018. Obtenido de <http://www.trabajo.gob.ec/ministerio-del-trabajo-establece-salario-basico-unificado-2018/>
- Mokate, K. (Junio de 1999). *Cepal*. Obtenido de Eficacia, eficiencia, equidad y sostenibilidad: ¿Qué queremos decir?: https://www.cepal.org/ilpes/noticias/paginas/9/37779/gover_2006_03_eficacia_eficiencia.pdf
- Muñoz Arias, M. J. (2018). *Proyecto Beca Socioeconómica*.
- Nishizawa, T. (25 de Octubre de 2002). Alfred Marshall on Human Capital and Future Generations. *Hitotsubashi University Repository*, 53(4), 305-321. doi:10.15057/21760
- OEA. (2011). *Desigualdad en Inclusión Social en las Américas: 14 ensayos*. Organización de los Estados Americanos. doi:<https://www.oas.org/docs/desigualdad/libro-desigualdad.pdf>
- Organización Mundial de la Salud. (2020). *Additional Background Material on the Draft Comprehensive Mental Health Action Plan 2013-2020*. Obtenido de https://www.who.int/mental_health/mhgap/background_material_action_plan_2013_2020.pdf?ua=1#:~:text=2012%2F9789241548410_eng.pdf-

,Psychosocial%20disabilities,inclusing%20stigma%2C%20discrimination%20and%20exclusion

Peña, D. (2002). *Análisis de datos multivariantes*. España: McGraw-Hill Interamericana de España, S.L.

Pérez, H. E., Laguado, A. C., & Martínez, A. (s.f.). *bdigital. Repositorio Institucional UN*. (O. N. Planeación, Ed.) Obtenido de Perfil socioeconómico de los estudiantes de la Universidad Nacional:
http://www.bdigital.unal.edu.co/3670/9/10_Perfil_socioeconómico_de_los_estudiantes_de_la_Universidad_Nacional_de_Colombia_2_2A.PDF

Pontificia Universidad Católica del Perú. (2008). *Reglamento General del Sistema de Becas y Crédito Educativo*. Perú. Obtenido de <https://s3.amazonaws.com/files.pucp.edu.pe/posgrado/wp-content/uploads/2014/09/01170529/reglamentogeneraldeelsistemadebecasycreditoeducativo.pdf>

Pontificia Universidad Javeriana. (01 de Septiembre de 2014). *Lineamientos Generales de las Becas y Estímulos a Estudiantes*. Bogotá, Colombia. Obtenido de <https://www.javeriana.edu.co/documents/10179/48161/Lineamientos+Generales+Becas+y+Est%C3%ADmulos+a+Estudiantes+Pontificia+Universidad+Javeriana+-+Sede+Central.pdf/9fd98f0d-a449-4b7b-bf02-b0ec4c16f149>

PUCE. (2017). *Modelo Educativo de la Pontificia Univesidad Católica del Ecuador*. Obtenido de <https://www.puce.edu.ec/intranet/documentos/Reglamentos/PUCE-Modelo-Educativo-06-03-2017.pdf>

PUCE. (2017). *Reglamento General de Estudiantes*. Recuperado el 28 de Mayo de 2019, de <https://www.puce.edu.ec/intranet/documentos/Reglamentos/PUCE-Reglamento-General-de-Estudiantes.pdf>

PUCE. (2018). *Documentos para asignación de la beca socio-económica*. Obtenido de Dirección de Beneficios Económicos: <https://www.puce.edu.ec/documentos/dge/PUCE-Requisitos-Beca-Socioeconomica-2018.pdf>

PUCE. (s.f.). *Pontificia Universidad Católica del Ecuador*. Obtenido de Dirección de Beneficios Económicos: <https://www.puce.edu.ec/beneficios-economicos.php>

Qian, Y. (2006). *K-means algorithm and its application for clustering companies listed in Zhejiang province*. School of Finance, Zhejiang University of Finance & Economics.

Regueyra Edelman, M. G. (15 de Noviembre de 2010). El Sistema de Becas por Condicion Socioeconómica y la Accesibilidad a la Educación Universitaria: Una Mirada desde los Indicadores de Gestión 2000-2008. *Revista Electrónica "Actualidades Investigativas en Educación"*, 10, 1-32. Obtenido de <http://www.redalyc.org/articulo.oa?id=44717988008>

Schultz, T. W. (1985). *Invirtiendo en la Gente*. Editorial Ariel, S.A.

Stiglitz, J. E. (2000). *La Economía del Sector Público, 3era edición*. Antoni Bosch.

Szirmai, A. (2010). *Dynamics of socio-economic development: an introduction. 2nd ed.* . Cambridge University Press.

Tan, P.-N., Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining*. Pearson Education.

- Tierney, W. G., & Hentschke, G. C. (2007). New players, different game: Understanding the rise of for-profit colleges and universities. 2-9.
- Tisdell, C. (Marzo de 1984). Slack and Strain in Efficient Budgeting and Resource Allocation by Organizations. *Managerial and Decision Economics*, 5(1), 54-57. Recuperado el 13 de Abril de 2019, de <http://www.jstor.org/stable/2487657>
- Universidad Católica de Córdoba. (08 de Agosto de 2018). Digesto-UCC. *Resolución Relectoral N° 1995*. Córdoba, Argentina. Obtenido de <https://www.ucc.edu.ar/archivos/documentos/Institucional/DIGESTO/Vicerrectorado%20Academico/Secretaria%20Academica/RR-BECAS-UCC-2018.pdf>
- Universidad Centroamericana José Simeón Cañas. (06 de Abril de 2018). *Reglamento del Programa de Becas Parciales de la Universidad Centroamericana José Simeón Cañas*. El Salvador. Obtenido de <http://www.uca.edu.sv/wp-content/uploads/2018/04/reglamento-programa-becas-parciales.pdf>
- Universidad Iberoamericana Ciudad de México. (Agosto de 2013). Reglamento de Becas y Financiamiento Educativo Educativo para Licenciatura. 37-50. México. Obtenido de <https://ibero.mx/sites/all/themes/ibero/descargables/alumnos/becas.pdf>
- Villanueva, J. (2018). Ventajas "Comparativas o Ventajas "Competitivas". *Boletín de Lecturas Sociales y Económicas*, 41-49. Obtenido de <http://200.16.86.50/digital/33/revistas/blse/villanueva18-18.pdf>

Anexos

Anexo n. ° 1

Cuadro 1: Estandarización de rangos de proporción de gasto

Rango por decil en función de la muestra	Rango
promedio decil 10 <gasto< máximo decil 10	100
promedio decil 9 <gasto< máximo decil 9	90
promedio decil 8 <gasto< máximo decil 8	80
promedio decil 7 <gasto< máximo decil 7	70
promedio decil 6 <gasto< máximo decil 6	60
promedio decil 5 <gasto< máximo decil 5	50
promedio decil 4 <gasto< máximo decil 4	40
promedio decil 3 <gasto< máximo decil 3	30
promedio decil 2 <gasto< máximo decil 2	20
promedio decil 1 <gasto< máximo decil 1	10
Si la proporción supera al valor máximo del decil 10	ALERTA (999)

Fuente y elaboración: Dirección de Beneficios Económicos (2018).

Anexo n. ° 2

Cuadro 2: Asignación de categoría según volumen de ventas e ingreso per cápita

Volumen de ventas	Ingreso per cápita	Habilitado
1-9999	Ingreso per cápita mínimo 89 USD	Habilitado para correr metodología SBS
	Ingreso per cápita máximo 250 USD	
10.000-20.000	Ingreso per cápita mínimo 250 USD	Habilitado para correr metodología SBS
	Ingreso per cápita máximo 500 USD	
20.000-30.000	Ingreso per cápita mínimo 500 USD	Asignar categoría F
	Ingreso per cápita máximo 750 USD	
30.000-50.000	Ingreso per cápita mínimo 750 USD	Asignar categoría F
	Ingreso per cápita máximo 1250 USD	
50.000-100.000	Ingreso per cápita mínimo 1250 USD	Asignar categoría F
	Ingreso per cápita máximo 2500 USD	
Más de 100.000	Ingreso mínimo más de 2500 USD	Asignar categoría F

Fuente y elaboración: Dirección de Beneficios Económicos (2018).

Anexo n. ° 3

Cuadro 3: Estandarización de los componentes de vulnerabilidad

Subdimensión asociada a Vulnerabilidad	Característica	Estandarización	
		Cumple característica	No cumple característica
Ubicación Geográfica	Estudiantes que residen fuera de la provincia de Pichincha.	100	0

Ocupación	Si el familiar trabaja por cuenta propia o es jornalero.	100	0
Desempleo	Si el familiar ha sufrido procesos de desempleo	100	0
Tenencia de la vivienda	Si no hay propiedad sobre la vivienda por hipotecada o arrendamiento	100	0
Acceso a Servicios Públicos de Agua, luz, alcantarillado	Si la vivienda No recibe alumbrado o agua por red pública y adentro de su casa	100	0
Bienes y Servicios	No tiene Internet	100	0
Crédito Educativo	Si el estudiante cuenta con un crédito educativo.	100	0
Discapacidad	Si uno de los miembros de la familia es discapacitado	100	0
Tipo de Colegio	Si el estudiante proviene de un colegio fiscal o municipal.	100	0
Endeudamiento Corriente (tarjetas de crédito)		100	0

Fuente y elaboración: Dirección de Beneficios Económicos (2018).

Anexo n. °4

Código de programación de los modelos alternativos en R

#####clasificación con variables cualitativas y cuantitativas#####

```
setwd("C:/Users/Cami/Desktop/datostesis")
```

```
data <- read.csv("data.csv", sep = ";")
```

```
nrow(data)
```

```
View(data)
```

```
colnames(data)
```

```
#NA
```

```
temp1<-is.na(data)
```

```
as.numeric(temp1)
```

```
sum(temp1)
```

```

#No existen NAs

#Se quita variable de categoría final
ndata <- data[-1]

#Prueba F: para determinar el número de centroides/grupos
n <- nrow(ndata)
SCDG <- n-1*sum(apply(ndata,2,var))

for (i in 1:10) {
  SCDG[i] <- sum(kmeans(ndata,i,nstart = 100)$withinss)
}

(SCDG[2]-SCDG[3])/(SCDG[3]/n-2-1)
(SCDG[3]-SCDG[4])/(SCDG[4]/n-3-1)
(SCDG[4]-SCDG[5])/(SCDG[5]/n-4-1)
(SCDG[5]-SCDG[6])/(SCDG[6]/n-5-1)
(SCDG[6]-SCDG[7])/(SCDG[7]/n-6-1)
(SCDG[7]-SCDG[8])/(SCDG[8]/n-7-1)
(SCDG[8]-SCDG[9])/(SCDG[9]/n-8-1)
(SCDG[9]-SCDG[10])/(SCDG[10]/n-9-1)
(SCDG[10]-SCDG[11])/(SCDG[11]/n-10-1)
(SCDG[11]-SCDG[12])/(SCDG[12]/n-11-1)
(SCDG[12]-SCDG[13])/(SCDG[13]/n-12-1)
(SCDG[13]-SCDG[14])/(SCDG[14]/n-13-1)
(SCDG[14]-SCDG[15])/(SCDG[15]/n-14-1)
(SCDG[15]-SCDG[16])/(SCDG[16]/n-15-1)
(SCDG[16]-SCDG[17])/(SCDG[17]/n-16-1)
(SCDG[17]-SCDG[18])/(SCDG[18]/n-17-1)
(SCDG[18]-SCDG[19])/(SCDG[19]/n-18-1)
(SCDG[19]-SCDG[20])/(SCDG[20]/n-19-1)

```

```
(SCDG[20]-SCDG[21])/(SCDG[21]/n-20-1)
```

```
#Clusters
```

```
set.seed(123)
```

```
kmeans.est <- kmeans(ndata,centers = 5, nstart = 120)
```

```
kmeans.est
```

```
#Between_SS/ Total_SS: mientras más cercano a 100% es mejor la medida de calidad
```

```
kmeans.est$cluster
```

```
#Medias
```

```
kmeans.est$centers
```

```
#Tamaño de los clusters: cuantos elementos se encuentran en cada clúster
```

```
kmeans.est$size
```

```
#Número de iteraciones
```

```
kmeans.est$iter
```

```
#Comparación de resultados
```

```
o <- order(kmeans.est$cluster)
```

```
resultados <- data.frame(data$cat_final[o],kmeans.est$cluster[o])
```

```
write.table(resultados[-1],file = "res_vm.csv",row.names = F)
```

```
write.table(resultados,file = "res_vm1.csv",row.names = F)
```

```
medias <- data.frame(kmeans.est$centers)
```

```
write.table(medias,file = "medias.csv",row.names = F)
```

```
#####clasificación con variables cuantitativas#####
```

```
setwd("C:/Users/Cami/Desktop/datostesis")
```

```
data <- read.csv("datac.csv", sep = ";")
```

```
nrow(data)
```

```
View(data)
```

```

colnames(data)

#NA
temp1<-is.na(data)
as.numeric(temp1)
sum(temp1)
#No existen NAs

#Se quita variable de categoría final
ndata <- data[-1]

#Clusters
set.seed(123)
kmeans.estc <- kmeans(ndata,centers = 5, nstart = 120)
kmeans.estc
#Between_SS/ Total_SS: mientras más cercano a 100% es mejor la medida de calidad
kmeans.estc$cluster
#Medias
kmeans.estc$centers
#Tamaño de los clusters: cuantos elementos se encuentran en cada clúster
kmeans.estc$size
#Número de iteraciones
kmeans.estc$iter

#Comparación de resultados
o <- order(kmeans.estc$cluster)
resultados <- data.frame(data$cat_final[o],kmeans.estc$cluster[o])
write.table(resultados[-1],file = "res_vc.csv",row.names = F)
write.table(resultados,file = "res_vc1.csv",row.names = F)

medias <- data.frame(kmeans.estc$centers)

```

write.table(medias,file = "mediasc.csv",row.names = F)

Anexo n. ° 5

Cuadro 4: Tablas de aranceles: estudiantes carreras nuevas y rediseñadas PUCE-Q (2018)

		A	B	C	D	E
CIENCIAS SOCIALES Y HUMANIDADES	ARANCEL TIPO 1	1900	2120	2505	2890	3850
	ARANCEL TIPO 3	950	1060	1253	1445	1925
	ARANCEL TIPO 4	475	530	626	723	963

CIENCIAS DE LA SALUD	ARANCEL TIPO 1	1900	2200	2600	3000	4000
	ARANCEL TIPO 3	950	1100	1300	1500	2000
	ARANCEL TIPO 4	475	550	650	750	1000

CIENCIAS NATURALES E INGENIERIAS	ARANCEL TIPO 1	2075	2285	2700	3115	4150
	ARANCEL TIPO 3	1038	1143	1350	1558	2075
	ARANCEL TIPO 4	519	571	675	779	1038

COMUNICACIÓN Y ARTE Y DISEÑO	ARANCEL TIPO 1	2125	2340	2765	3190	4250
	ARANCEL TIPO 3	1063	1170	1383	1595	2125
	ARANCEL TIPO 4	531	585	691	798	1063

ARQUITECTURA	ARANCEL TIPO 1	2400	2700	3100	3700	4750
	ARANCEL TIPO 3	1200	1350	1550	1850	2375
	ARANCEL TIPO 4	600	675	775	925	1188

MEDICINA	ARANCEL TIPO 1	3000	3300	3900	4500	5589
	ARANCEL TIPO 3	1500	1650	1950	2250	3000
	ARANCEL TIPO 4	750	825	975	1125	1500

RANGOS	
ARANCEL TIPO 1	DE 16 EN ADELANTE
ARANCEL TIPO 2	DE 9 A 15 CRÉDITOS
ARANCEL TIPO 3	DE 1 A 8 CRÉDITOS