

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ciencias Sociales y Humanísticas

Economía Regional: Modelo predictivo y análisis de variables
influyentes sobre la Inversión de Obras Públicas de GAD
cantonales del Ecuador

PROYECTO INTEGRADOR

Previo la obtención del Título de:

Economista con Mención en Gestión Empresarial

Presentado por:

Xavier Andrés Tobar Escobar

Ariana Janeth Miñaca Toro

GUAYAQUIL - ECUADOR

Año: 2021

DEDICATORIA

Este trabajo de investigación está dedicado a mi familia, quienes han apoyado amablemente hasta que mi investigación estuvo completamente terminada, y a mis compañeros más allegados, durante los últimos meses, quienes me han alentado atentamente con su más completa y verdadera atención para realizar mi trabajo con la verdad.

Xavier Andrés Tobar Escobar

El presente trabajo va dedicado con infinito amor a mis padres por estar para mí en todo momento, a mis abuelitos que con su sabiduría guían mi camino; de forma especial para quienes ya no están físicamente conmigo, Arturo Miñaca Poveda (+) y Lidia Arévalo Benavides (+), me enseñaron a no rendirme porque el estudio es la herencia más importante en la vida. A mis hermanos, por su cariño y fortaleza todo el tiempo. No puedo olvidar a mis familiares y amigos más cercanos, por las palabras de motivación para alcanzar mi meta.

Ariana Janeth Miñaca Toro

AGRADECIMIENTOS

A lo largo de la redacción de este proyecto, hemos recibido una gran cantidad de apoyo y asistencia. En primer lugar, queremos agradecer a nuestro tutor Xavier Villavicencio cuyo aporte fue considerable en la obtención de documentos clave para el proyecto. Por otro lado, agradecemos al tutor Yeltsin Castro Loaiza por su invaluable experiencia y apoyo con recomendaciones para agregar a la cimentación de la metodología.

Xavier Tobar y Ariana Miñaca Toro

Agradezco a Dios, por la vida y sus bendiciones. A mis padres, hermanas, familiares y amigos por darme su apoyo incondicional. A mi compañera de tesis, Ariana. A su vez me gustaría agradecer al Dr. Gonzalo Sánchez por inspirarme a adentrarme más en los temas de la econometría y la investigación económica.

Xavier Tobar Escobar

Agradezco a Dios, por la vida y sus bendiciones. A mis padres, hermanos, familiares y amigos por creer en mí y darme su apoyo incondicional. A mi amigo y compañero de tesis Xavier, por su confianza y optimismo.

Ariana Miñaca Toro

DECLARACIÓN EXPRESA

“Los derechos de titularidad y explotación, nos corresponde conforme al reglamento de propiedad intelectual de la institución; *Xavier Andrés Tobar Escobar* y *Ariana Janeth Miñaca Toro* damos nuestro consentimiento para que la ESPOl realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual”

Xavier Tobar E

Xavier Andrés Tobar
Escobar



Ariana Janeth Miñaca
Toro

EVALUADORES

.....
Juan Carlos Campuzano Sotomayor
PROFESOR DE LA MATERIA

A handwritten signature in blue ink on a light-colored background. The signature is cursive and appears to read 'Xavier Alfredo Villavicencio Córdova'.

Xavier Alfredo Villavicencio Córdova
PROFESOR TUTOR

RESUMEN

El objetivo de esta investigación es analizar los factores que influyen en la inversión de obras públicas de los 134 GAD Municipales del Ecuador en el año 2018 cuyos datos estaban disponibles, utilizando el método estadístico de clasificador de árbol que permite la identificación de las relaciones entre sus variables. Este modelo se basará en información obtenida de encuestas como la ENEMDU y GIRS; a su vez se utilizará un informe del ministerio de finanzas para obtener los datos de inversión en obras públicas y costos de salarios. Se aplica un criterio de selección de variables utilizando un modelo de XGBoost para obtener las variables que más aporten al modelo tanto en cantidad de información que ofrecen como su aportación a la clasificación. Se realizaron 3 modelos de árbol para contrastar su desempeño, pero solo los dos árboles basados en el modelo CART de la librería rpart de R se obtuvieron valores de clasificación satisfactorios. Las variables más significativas fueron los costos en salarios y el número de empleados, puesto que si exceden el 50% y 191 respectivamente los municipios tienden a invertir menos en obras.

Palabras Clave: GAD Municipal, Inversión, Salarios, Machine learning, Árbol Clasificador.

ABSTRACT

The aim of this research is to analyze the factors that influence the investment in public works of the 134 Municipal GADs of Ecuador in 2018 whose data were available, using the statistical method of tree classifier that allows the identification of the relationships between these variables. This model will be based on information obtained from surveys such as the ENEMDU and GIRS; in turn, a report from the finance ministry will be used to obtain data on investment in public works and salary costs. A variable selection criterion is applied using an XGBoost model to obtain the variables that contribute the most to the model, both in terms of the amount of information they offer and their contribution to the classification. Three tree models were made to contrast their performance, but only the two trees based on the CART model of the R rpart library obtained satisfactory classification values. The most significant variables were costs in salaries and the number of employees, since if they exceed 50% and 191 respectively, municipalities tend to invest less in public works.

Keywords: *Municipal GADs, Investment, Salaries, Machine learning, Decision trees.*

ÍNDICE

| | |
|---|-----|
| RESUMEN | I |
| ABSTRACT | II |
| ABREVIATURAS | V |
| ÍNDICE DE FIGURAS..... | VI |
| ÍNDICE DE TABLAS | VII |
| CAPÍTULO I | 1 |
| 1. Introducción..... | 1 |
| 1.1 Descripción del problema | 3 |
| 1.2 Justificación del problema..... | 4 |
| 1.3 Pregunta de Investigación | 5 |
| 1.4 Objetivos..... | 6 |
| 1.4.1 Objetivo General..... | 6 |
| 1.4.2 Objetivos Específicos | 6 |
| 1.5 Variable de interés | 6 |
| 1.6 Marco teórico..... | 7 |
| 1.6.1 Literatura Principal..... | 7 |
| 1.6.2 Administración Pública | 8 |
| 1.6.3 Factores Relevantes de la Gestión Pública..... | 10 |
| 1.6.4 Indicadores indirectos de desempeño municipal | 14 |
| 1.6.5 Literatura Secundaria | 16 |
| CAPITULO II | 18 |
| 2. Metodología..... | 18 |
| 2.1 Fuente de datos e información | 18 |
| 2.2 Tratamiento y depuración | 18 |
| 2.3 Descripción de las variables | 20 |
| 2.4 Operacionalización de las variables | 23 |
| 2.4.1 Construcción de indicadores y operación de variables..... | 23 |
| 2.4.2 Conceptualización de variables | 24 |
| 2.5 Metodología Propuesta | 26 |
| 2.5.1 Herramientas y criterios estadísticos..... | 26 |
| 2.5.2 XGBoost para selección de variables..... | 26 |
| 2.5.3 Índice de diversidad de Shannon..... | 27 |
| 2.5.4 Árbol de Clasificación y Regresión (CART)..... | 28 |

| | | |
|--------------------|--|----|
| 2.5.5 | Índice de GINI..... | 30 |
| 2.5.6 | Matriz de confusión | 30 |
| CAPITULO III | | 33 |
| 3. | Resultados | 33 |
| 3.1.1 | Análisis Exploratorio..... | 33 |
| 3.1.2 | Distribución de variables categóricas | 34 |
| 3.1.3 | Revisión de asimetría de variables categóricas | 37 |
| 3.1.4 | Matriz de correlaciones | 38 |
| 3.2 | Selección de variables (XGBoost)..... | 38 |
| 3.3 | Árbol de Clasificación..... | 43 |
| 3.3.1 | Interpretación de los árboles..... | 45 |
| CAPITULO IV..... | | 50 |
| 4. | Conclusiones y Recomendaciones | 50 |
| 5. | Bibliografía | 52 |

ABREVIATURAS

ESPOL Escuela Superior Politécnica del Litoral

GAD Gobierno Autónomo Descentralizado

COOTAD Código Orgánico de Organización Territorial, Autonomía y Descentralización

AME Asociación de Municipalidades del Ecuador

XGBOOST Extreme Gradient Boosting

CART Classification and Regression Trees

ÍNDICE DE FIGURAS

| | |
|---|----|
| Gráfico 1. Aplicación de la norma ecuatoriana | 9 |
| Gráfico 2. Transferencias por el Modelo de Equidad Territorial..... | 10 |
| Gráfico 3: cantidad de datos faltantes..... | 19 |
| Gráfico 4 Estructura de árbol | 29 |
| Gráfico 5 Matriz de confusión..... | 31 |
| Gráfico 6 Medidas Matriz de Confusión..... | 31 |
| Gráfico 7 top 20 municipios en costo de salarios | 33 |
| Gráfico 8 Boxplot 1 | 34 |
| Gráfico 9 Boxplot 2 | 35 |
| Gráfico 9 scatter plot 1..... | 36 |
| Gráfico 10 simetría de variables..... | 37 |
| Gráfico 11 Matriz de Correlaciones 1..... | 38 |
| Gráfico 12 Importancia de variables(weight) | 40 |
| Gráfico 12 Importancia de variables(Gain) | 40 |
| Gráfico 13 Matriz de Correlaciones 2..... | 41 |
| Gráfico 14 Árbol de decisiones CART(w/costo de salarios)..... | 46 |
| Gráfico 15 Árbol de Decisiones(W/o costo de salarios)..... | 48 |

ÍNDICE DE TABLAS

| | |
|---|----|
| Tabla 1 variable dependiente | 20 |
| Tabla 2 variables Independientes | 20 |
| Tabla 3 variables de control | 22 |
| Tabla 4 XGBoost número de variables | 39 |
| Tabla 5 variables seleccionadas | 42 |
| Tabla 6 Comparación de modelos | 44 |
| Tabla 7 resumen mejores modelos | 45 |

CAPÍTULO I

1. INTRODUCCIÓN

Según la Asociación de Municipalidades Ecuatorianas (AME), en el 2020, la deuda del Estado con los 221 municipios ascendió a USD 1.200 millones. En este periodo de crisis sanitaria, dicha deuda ha generado mayor malestar a los 152 GAD Municipales definidos como pequeños, debido a que estos han requerido mayores recursos para enfrentar los efectos de la pandemia. Cabe señalar que los escasos recursos por lo general en su mayoría han sido destinados a pagos de sueldos y salarios de funcionarios, lo que ha originado que la inversión en obra pública sea mínima.

El plan de descentralización propuesto tiene como objetivo superar las deficiencias de la asignación de recursos nacionales a los gobiernos locales, y se basa en un nuevo marco institucional responsable de promover, regular y transferencia y ejercicio de poderes exclusivos y concurrentes de la organización ejercidos previamente, incluidos los nuevos que ya han sido ejercitados se transfiera gradualmente (SENPLADES, 2012) (COOTAD, 2010)

Los municipios del país reciben fondos del presupuesto general del estado. Estos recursos se utilizan para obras públicas en diferentes provincias y ciudades. El 9% del presupuesto cada año se asigna a los municipios, pero la transferencia de estos recursos se ajustará. En el caso de menos recursos, cada municipio deberá priorizar el trabajo o tomar medidas para obtener más ingresos (COOTAD, 2010).

Esta situación conlleva replantear la manera en la que se consideran ciertos parámetros para la repartición de presupuesto e inclusive algunas de las directrices del COOTAD con respecto a que tanta libertad o falta de control debería tener un municipio. La falta de transparencia para el público general y el uso de datos herméticos para las decisiones que se toman respecto a la asignación de presupuestos, alientan al desarrollo de herramientas adicionales que permitan profundizar en las relaciones de las diferentes características de un municipio con datos más accesibles.

El presente documento desarrolla una propuesta metodológica para predecir si un GAD Municipal invertirá por encima de la media en obras públicas. Es decir, esta propuesta se basa en obtener una medición con datos accesibles que profundiza las características que tienen los municipios al invertir menos en obras públicas.

1.1 Descripción del problema

Se conoce que la mayoría de los municipios pequeños tienen un gran porcentaje de gastos en salarios de funcionarios públicos y a su vez un bajo porcentaje en inversiones de obras públicas (Guerra, 2020). El tema principal de esta investigación es la baja tasa de inversión en obras públicas y que factores influyen en la misma.

La primera causa importante en el desempeño de un municipio es la participación ciudadana, que se plantean en función de la capacitación universitaria que existe en los municipios y en general en los municipios pequeños existe mayor gasto en salario que en sus obras públicas. El nivel de estudios en la población adulta se utilizó como proxy de la participación ciudadana una investigación pasada realizada por De Borger y Kerstens. Se determinó que la Participación ciudadana tenía impacto sobre la eficiencia municipal (Frankle, 2007).

La segunda causa que se identificó que podría estar aportando al problema es la gestión pública desactualizada con la que se manejan la mayoría de los municipios en el país. Debido a que estos se manejan con datos de poca trazabilidad y transparencia, en donde cada municipio presenta sus datos de forma pública en diferentes formatos y pocas veces digitalizados. Esto incentiva a que no se realicen investigaciones en el ámbito privado, sino que sean realizadas por investigadores afiliados por el gobierno central y por lo tanto no permite cuestionar la eficiencia de sus gestiones.

A partir de estas causas se podría decir que se deriva otro factor que se traduciría como una falta de control o estandarización por parte del gobierno central que permita elevar el estándar de la gestión administrativa pública en los distintos municipios del Ecuador, que le permitiría al país dar un salto tecnológico en lo que concierne sus métodos y medidas administrativas debido a que serán tomadas basadas en datos transparentes.

1.2 Justificación del problema

Según la Secretaría Nacional de Planificación y Desarrollo, existen criterios para la asignación de recursos a los gobiernos autónomos descentralizados, como por ejemplo: tamaño y densidad poblacional, necesidades básicas de jerarquía insatisfechas, consideraciones relacionadas con la población permanente en el territorio de cada GAD, logros en la mejora de los niveles de vida, esfuerzos financieros y administrativos, y cumplimiento de metas nacionales (Plan de Desarrollo y Plan del gobierno autónomo descentralizado GAD) para garantizar la equidad en la asignación de recursos y la eficiencia de los destinos, los cuales responden a la provisión de bienes y servicios públicos al área local a la población de todos los niveles de gobierno (SENPLADES, 2012).

A pesar de las demandas por parte de los municipios pequeños por recibir el monto de presupuesto asignado, se plantea la pregunta de en qué invierten ese dinero y si es que las inversiones que ellos realizan realmente están brindando un beneficio a su comunidad. No se han realizado investigaciones públicas que profundicen más en este tema para determinar cuál podría ser una razón influyente en ese fenómeno debido a la dificultad de conseguir datos de cada municipio a nivel cantonal, por lo que en esta investigación se propone utilizar la variable binaria que mide si un municipio invierte por encima del promedio en obras públicas para intentar obtener un estimador que permita analizar las características más influyentes en la inversión de obras públicas de un municipio.

En un estudio anterior Vanden Eeckaut et al. (1993) y De Borger y Kerstens (1996a) pudieron determinar que la variable de participación ciudadana, la cual se aproximó por el nivel educativo de la población adulta, tiene un impacto positivo sobre el grado de eficiencia municipal (Frankle, 2007).

Para obtener datos que nos puedan ayudar a profundizar en este tema se utilizaran características sociodemográficas, ambientales y poblacionales. Por tanto,

con esta propuesta de investigación se plantea realizar un modelo estadístico para seleccionar características que puedan darnos indicios sobre qué municipio invertiría más en obras públicas y obtener reglas en las cuales esa diferencia se vuelve significativa.

A lo largo de las secciones previas a la investigación se han encontrado problemas con la fiabilidad de los datos y la ausencia de un registro continuo de datos relevantes a municipios, sobre todo esta problemática se ve aumentada en los municipios pequeños. Por lo que para este estudio se utilizarán datos limitados que se basen solamente en el 2018 que es en el año donde se encontró mayor cantidad de información necesaria para el estudio.

El modelo será basado en datos transversales y se utilizará árbol de decisiones por lo cual solo se podrá observar las relaciones que tienen estas variables con respecto al tema a estudiar mas no se podrán hacer inferencias causales. De igual manera se busca otorgar una herramienta adicional para predecir que municipios estarían más susceptibles a invertir por debajo de la media.

1.3 Pregunta de Investigación

¿Es la autonomía financiera la característica más influyente de un municipio en la inversión pública?

¿Qué otros factores o características tendrían un municipio que se traduzcan en una inversión pública por encima del promedio?

1.4 Objetivos

1.4.1 Objetivo General

Analizar los factores que influyen en la inversión de obras públicas de los 134 GAD Municipales del Ecuador en el año 2018 cuyos datos estaban disponibles, utilizando el método estadístico de clasificador de árbol que permite la identificación de las relaciones entre sus variables.

1.4.2 Objetivos Específicos

- Aplicar un criterio de selección de variables para obtener las de mayor importancia para la especificación del modelo.
- Utilizar el método estadístico de árbol de decisiones para la identificación de las variables que inciden en la probabilidad de inversión en obra pública por parte de los municipios.
- Evaluar en que punto cada una de las variables identificadas inciden en la inversión y su concentración de los datos en dicho nodo.

1.5 Variable de interés

La variable de estudio será una variable binaria que indica si un municipio invierte por encima de la media en obras públicas y las variables dependientes serán características poblacionales, ambientales y sociodemográficas como: capacitación universitaria por cantón, permisos de gestión de residuos, calles susceptibles a barrido, gasto en salarios, número de empleados públicos, etc. Se describirá con mayor profundidad las variables del estudio en la sección de metodología.

1.6 Marco teórico

1.6.1 Literatura Principal

El nuevo modelo de descentralización es en términos de unidad y equidad regional sus principios básicos, según lo estipulado en el artículo 272 de la Constitución y en el art. 192 del COOTAD existe estándares para la asignación de recursos a los gobiernos autónomos descentralizados, como por ejemplo: tamaño y densidad poblacional, necesidades básicas de jerarquía insatisfechas, consideraciones relacionadas con la población permanente en el territorio de cada gobierno autónomo disperso, logros en la mejora de los niveles de vida, esfuerzos financieros y administrativos, y cumplimiento de metas Plan Nacional de Desarrollo y Plan de Desarrollo del Gobierno Autónomo Descentralizado (GAD) para garantizar la equidad en la asignación de recursos y la eficiencia de los destinos del mismo, además garantizar la provisión de bienes y servicios públicos a la población de todos los niveles de gobierno (SENPLADES, 2012).

En la actualidad los 221 Gobiernos Autónomos Descentralizados Municipales del Ecuador, indudablemente tienen autonomía financiera, administrativa y política de acuerdo con el artículo 53 del Código Orgánico de Organización Territorial, Autonomía y Descentralización (COOTAD). Además, cuentan con la participación ciudadana; legislación y fiscalización, y ejecutiva de funciones y competencias de lo que establece el mismo (COOTAD, 2010).

1.6.2 Administración Pública

La Administración Pública tiene varios conceptos, la cual claramente es un elemento constitutivo al Estado. Al ser extensa su definición, se toma las palabras de Omar Guerrero, que define la administración pública como una entidad de material integral para la región y la gente, además es la unidad de muchas personas bajo un poder máximo (Guerrero, 2007). Primero, el concepto de administración pública se espera de la definición de naturaleza jurídica, es decir, de Constitución de la República del Ecuador y posteriormente de autores dedicados a su estudio. Así como está establecido en el artículo 227, la Administración Pública “constituye un servicio a la colectividad que se rige por los principios de eficacia, eficiencia, calidad, jerarquía, desconcentración, descentralización, coordinación, participación, planificación, transparencia y evaluación” (ECUADOR C. R., 2008).

La capacidad de instalación en todos los niveles del gobierno está aumentando. Las instituciones a través del Comité Nacional de Competencias, ya sea en términos de talentos, mejorar y automatizar mediante la aplicación de nuevos procesos tecnología para realizar mejoras y procesos que ayuden a brindar calidad, eficiencia y servicio de efectividad para los ciudadanos (Plan Nacional de Descentralización, 2013-2015).

Por otra parte, la administración pública, las entidades y agencias del gobierno central de Ecuador, y como esos regímenes desconcentrados y descentralizados, corresponden a observar cosas estables en la Constitución, leyes y sus respectivas disposiciones; actúan los departamentos de la administración pública. La misión es buscar una vida mejor para los ciudadanos y luego brindar productos y servicios de calidad. Bajo los principios de efectividad, eficiencia, transparencia y evaluación, la cotización es de alta calidad (SENPLADES, 2012).

Por lo tanto, la administración pública busca atender las diversas necesidades de la ciudadanía, orientar sus actividades de acuerdo con los principios de eficiencia

y eficacia, y considerar modelos de gestión eficaces para asegurar el aprendizaje y la mejora continua. En el siguiente gráfico, se observa cómo se aplica la norma de acuerdo a la supremacía de ley establecida en la Constitución de la República del Ecuador.

Gráfico 1. Aplicación de la norma ecuatoriana



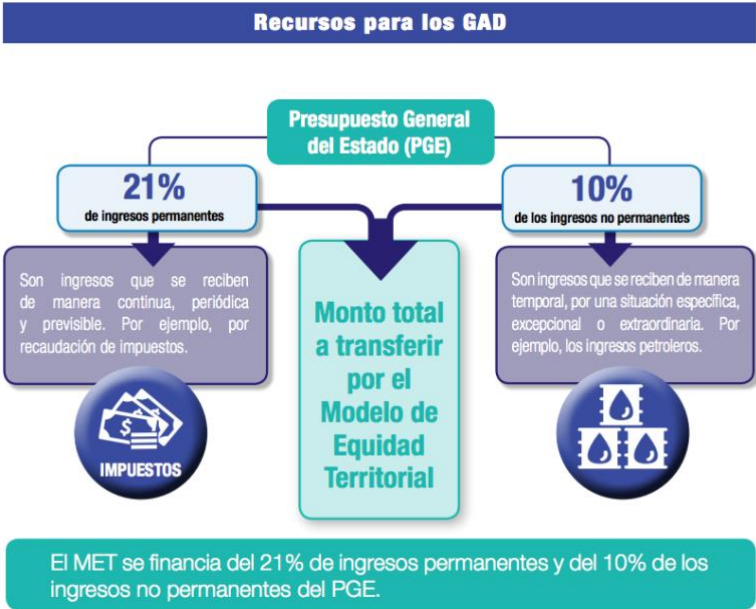
Elaboración: Imagen obtenida de la Constitución de la República del Ecuador, art. 425 (p.189)

Con base en el sustento legal y la aplicación del Código en el territorio de Ecuador, el marco regulatorio y legal en el que se basa esta investigación se detalla a continuación:

- Constitución del Ecuador
- Código Orgánico de Planificación y Finanzas Públicas
- Código Orgánico de Organización Territorial, Autonomía y Descentralización COOTAD.

Por otra parte, la asignación de recursos es un factor importante para el desarrollo de los Gobiernos Autónomos Descentralizados Municipales del Ecuador, en la situación actual, los GAD Municipales realizan varias competencias según el artículo 55 del Código Orgánico de Organización Territorial Autonomía y Descentralización (COOTAD), una de estas es la planificación del uso del sueldo para el gasto en obras pública (COOTAD, 2010).

Gráfico 2. Transferencias por el Modelo de Equidad Territorial



Elaboración: Imagen obtenida de Planificación, COOTAD artículo 192 (Gad, 2015)

1.6.3 Factores Relevantes de la Gestión Pública

A pesar del propósito que tienen las medidas legislativas antes mencionadas, según un estudio pasado sobre la gestión de los ingresos y su incidencia en la inversión pública del GAD de Salinas; el cual analizó el sector público de Salinas de los años 2010-2016. Se encontró que la gestión de los ingresos de dicho GAD no permite cubrir las necesidades básicas del cantón. Ya que del 100% de los ingresos

efectivos generados durante el periodo de 2010-2016 solo se destinó el 15,6% para inversión pública. (Margarita Panchana P., 2017)

Dentro del estudio de Panchana, se da importancia a la gestión administrativa de los ingresos y gastos del sector público, debido a que es pertinente que se desarrollen estrategias aplicadas y se tenga una disciplina fiscal. En otras palabras; los gerentes, directivos y todo servidor público deben ser profesionales capaces y fiel observadores de la ley normativa vigente. Por otro lado, Panchana concluye que el presupuesto de ingresos por lo general es liquidado anualmente con cartera vencida y como tal no se formulan ni aplican políticas idóneas de la gestión de cobro.

En los párrafos anteriores se habló sobre la importancia de la gestión administrativa de los ingresos, por lo que es pertinente considerar lo que influye en una buena gestión administrativa y este sería el factor de la capacitación de los funcionarios públicos. En un estudio exploratorio pasado realizado por Aragón y Casas, se buscó observar el efecto que tendría la carencia de capacidad técnica en el desempeño de los gobiernos locales del Perú. Los autores parten de la intuición de que la distribución de presupuestos no es la única limitante importante en el sector público y buscan profundizar más en el aspecto de las capacidades de gestión.

Dentro de la investigación de Aragón y Casas, se evaluó el impacto de la capacidad técnica sobre el desempeño de los municipios peruanos. En la investigación lograron capturar el efecto estimado de la propensión a gastar recursos sobre los gastos corrientes y de capital. Los datos que analizaron sugieren que la baja inversión ocurre mayormente en gastos de capital que en gastos corrientes. Las estimaciones que Aragón y Casa realizaron cubren el periodo de 2000-2006 y ha incluido casi todo el distrito de los municipios peruanos para esos años. Durante esos años encontraron un aumento en ingresos municipales de casi 300% los cuales eran proveniente de transferencias. La intuición sugiere que la falta de capacidad técnica afecta a la propensión a gastar en los gastos de capital y se sustenta con los datos que lograron recopilar.

A su vez también tomaron en cuenta otros factores como la población y el índice de pobreza. En donde se encontró que los distritos más pobres gastan más que el distrito promedio y los distritos con más población tienen una menor propensión a gastar. Según Aragón y Casas, esto sugiere que existe un gran número de necesidades básicas en los distritos más pobres por lo que podría haber más presión popular o consenso en la jurisdicción sobre la relevancia del gasto de capital y los municipios tienden a ser más proactivos en términos de ejecución del capital.

En el caso de la población encontraron que los distritos con mayor número de habitantes tienden a gastar menos. Ese resultado podría implicar un problema de escalas, en donde a mayor población la inversión de los proyectos exige más recursos y habilidades técnicas. Si los distritos donde la población es grande (probablemente los urbanos) tienen una menor propensión a gastar, la actitud de los votantes será menor. Adverso a los gobiernos locales y apoyará a las críticas del gobierno central. Este resultado puede explicar la revocación de un número importante de alcaldes principalmente en áreas urbanas. Según Aragón y Casas: "Los votantes sienten que hay una gran cantidad de recursos en las cuentas de los municipios, pero los alcaldes y funcionarios son incapaces de transformar estos recursos en inversión de proyectos" (Casas, 2008)

El mayor hallazgo de la investigación realizada por los autores reside en la relación que se encontró entre la propensión a gastar y la falta de capacidad técnica en manejo de proyectos. Con esto los autores sugieren la importancia que se debería prestar no solo a una educación tradicional para los cargos de gestión pública sino también un entrenamiento dentro del campo laboral para poder desarrollar habilidades como la gestión de proyectos.

Por otro lado, En 2010, Carlos Haefner, en su estudio "Propuesta para una política de desarrollo municipal de calidad en Chile", busca como mejorar la calidad de la gestión de los servicios municipales y la satisfacción de los usuarios internos y externos, se concluye que, a través del modelo de gestión municipal, se avanza en su

gestión. Este es el proceso de auto observación y auto examen del municipio y sus funcionarios sobre su posición en los asuntos de gestión. Significa que se debe identificar metas relacionadas con el modelo y asumir un plan de mejora que "desencadena" todo el proceso de mediano a largo plazo hasta alcanzar la mejor situación de certificación.

La aplicación del sistema de gestión es un mecanismo de evaluación externa y reconocimiento público. En comparación con la práctica de gestión del modelo, el municipio ha alcanzado un cierto nivel de gestión municipal. Para lograr las metas anteriores, el sistema de certificación de servicios municipales busca incentivar a los municipios a iniciar y mantener un proceso permanente y continuo de mejoramiento de la gestión y los resultados hasta alcanzar un nivel de excelencia, es decir, deben brindar la capacidad de brindar calidad. servicios a sus usuarios.

Teniendo en cuenta las investigaciones que se han mencionado, se evidencia la necesidad de obtener un modelo que permita observar que variables influirían en el desempeño de un municipio y de tal manera otorgar una herramienta que asista a la formulación de estrategias para el beneficio de la comunidad.

Debido a que el objetivo de esta investigación es observar las variables que influyen en la inversión de obras públicas de un municipio y utilizar un modelo estadístico de árbol de decisiones para profundizar en sus significancias, se utilizarán investigaciones pasadas en donde se han abordado los temas de eficiencia y desempeño municipal.

Estas investigaciones coinciden siempre en el mismo problema el cual es la falta de información directa que indique los niveles de gestión pública. Según Porto, toda investigación aplicada tendrá problemas que se originan en la falta de información, pero utilizando información disponible se pueden llegar a resultados contrastables para futuras investigaciones. (Alberto Porto, Performance and efficiency measures in the public sector. The case of Bolivia, 2018). Debido a esta razón que se origina de la complejidad del sector público, se recomiendan utilizar indicadores indirectos que puedan dar indicios en el desempeño de dicho municipio o administración pública.

1.6.4 Indicadores indirectos de desempeño municipal

Teniendo en cuenta lo mencionado anteriormente y debido a la complejidad de obtener métricas de desempeño del sector público; Para esta investigación se utilizarán indicadores indirectos por medio de los servicios provistos de cada municipio. Según Porto tales indicadores podrían ser datos pertenecientes a sectores como: Ambiental, salud, educación, etc. (Alberto Porto, Performance and efficiency measures in the public sector. The case of Bolivia, 2018).

La intuición de esta selección radica en que el sector público tiene como objetivo proveer de beneficios a la comunidad de la cual están encargados. Entre los indicadores que se utilizaran en la investigación están los siguientes:

Ambientales:

- Longitud de calles susceptibles a barrido por área urbana y rural
- Capacitación y ordenanza de manejo de residuos
- Agua potable
- Sitios de depósito.

Estas variables ambientales permiten observar las características de cada municipio respecto a su gestión y enfoque ya sea si dicho municipio tiene mayor

enfoque en el sector urbano o rural. A su vez, la capacitación y ordenanza permite observar la participación que podría llegar a tener el municipio en función del bienestar de su comunidad.

Poblacionales y gestión:

- Autonomía financiera
- Número de empleados públicos
- Cantidad de dinero autofinanciado por empleado
- Costo de salarios públicos
- Inversión en obras publicas
- Población universitaria

En estas variables de población y gestión se encuentra la variable objetivo y otras que ayudarían a controlar el modelo. Se buscará encontrar la relación de estas categorías con la variable de inversión en obras publicas la cual se trabajará para realizar un modelo clasificador. La explicación sobre cada variable y su preparación se hablarán en la sección de metodología.

1.6.5 Literatura Secundaria

La literatura secundaria ayuda a la discusión, argumentación, ejemplificación o aplicación del marco teórico desarrollado previamente.

Como referencias secundarias se utilizará un trabajo anterior realizado por Josip mesaric y Dario Sebalj, el cual utiliza un árbol de decisiones para predecir el éxito académico de estudiantes. En este trabajo se comparan distintos métodos de árbol y al se concluye que el método de CART fue el más apropiado con un mejor porcentaje de clasificaciones correctas. “En la base de datos consiste de 450 estudiantes que se enrolaron en el curso de información de sistemas. Se usaron arboles de decisión y regresión logística. Dentro de los métodos de árboles de decisión está método CART, el cual fue el más exitoso con una predicción del 60.5% de clasificación correctas. (Josip Mesari, 2016)

A su vez se “podará” el árbol para reducir su complejidad y obtener un estimador que sea utilizable fuera de la muestra; Utilizando el método de Minimal cost complexity pruning (mccp). Este algoritmo está parametrizado por α (≥ 0) conocido como parámetro de complejidad. El parámetro de complejidad se utiliza para definir la medida de costo-complejidad, $R_\alpha(T)$ de un árbol dado T : $R_\alpha(T) = R(T) + \alpha |T|$ donde $|T|$ es el número de nodos terminales en T y $R(T)$ se define tradicionalmente como la tasa total de clasificación errónea de los nodos terminales.

Por otro lado para proporcionar algunas recomendaciones respecto al cumplimiento de la transparencia de la información divulgada por el departamento administrativo de parte del gobierno, hay algunas sugerencias para verificar la autenticidad de la información del municipio a partir de indicadores, incluida la calidad y transparencia de la información y las comunicaciones publicadas a través de las páginas web del municipio: una comparación de casos entre Ecuador y España "Narcisa Medranda En (2017), el mismo De acuerdo al marco legal implementado en LOTAIP en 2014, a través de los indicadores del método Mapa Infoparticipa, el nivel de credibilidad que existe en las páginas web municipales de Ecuador y España. La información de propiedad de la ciudadanía es reconocida en los GADS municipales de cada región de esta manera, y es obligación de informar y establecer una democracia participativa y efectiva entre ciudadanos y gobernadores.

Sobre cómo gestionar la información pública desde la calidad y la transparencia dentro de los municipios, puede encontrar un artículo en su sitio web, Transparencia en los municipios ecuatorianos: métodos y resultados, de Narcisa Medranda y Pedro Molina. (2018), mediante un análisis en profundidad del recorrido histórico de la transparencia, LOTAIP y la implementación del mapeo interactivo creado en España denominado Infoparticipa, conoció los resultados de la evaluación de la transparencia de la información insertada en varias páginas web municipales en Ecuador. Del mismo modo, están probados debido a la información incompleta sobre gestión de recursos y administración pública en varias páginas web, la transparencia es insuficiente.

CAPITULO II

2. METODOLOGÍA

2.1 Fuente de datos e información

Las fuentes de información que se han utilizado para esta investigación son las siguientes:

- De la encuesta del ENEMDU 2019, se obtuvo la información de las proyecciones poblacionales y población universitaria. Debido a que son encuestas a nivel de individuo, se utilizó el factor de expansión indicado por la ENEMDU para obtener estimaciones poblacionales.
- De la encuesta GIRS para los municipios cantonales se obtuvo los indicadores ambientales y las características de cada municipio respecto a su participación en el sector ambiental. Debido a que los datos eran a nivel de municipio, no se tuvo que aplicar ningún factor de expansión.
- Los datos de la variable de interés y otros posibles datos de gestión pública se obtuvieron de un informe realizado por el Ministerio de finanzas en el cual se detallaban indicadores de autonomía financiera, costos de salarios, entre otros.

2.2 Tratamiento y depuración

Para obtener una base de datos que se ajuste a las necesidades de la investigación era necesario llevar todos los datos al mismo nivel de observación de nuestra variable de interés. En este caso el data set que se utilizó como base fueron los datos que se encontraban en el informe del ministerio de finanzas el cual tenía observaciones a nivel de municipios cantonales.

Debido a que los datos de la encuesta de la ENEMDU estaban a nivel de individuo primero se procedió a obtener las variables de interés de esa encuesta utilizando el archivo de la base de datos disponible en la INEC. Luego de agrupar los

datos en función de las variables que se quieren analizar, se utilizaba el factor de expansión para obtener las proyecciones poblaciones de las variables que se obtuvieron.

Después de transformar las bases de datos a nivel cantonal, se procedió a utilizar un diccionario de variables en los cuales tenían los códigos de los cantones para poder realizar un merge entre estas tres bases de datos, utilizando la columna de cantones. Después del merge se obtuvo toda la información en la cual tenían intersecciones con la columna de cantones y se obtuvieron 25 variables para 134 cantones en la base de datos final.

Se decidió no reemplazar los datos faltantes y dejarlos debido a que el modelo de árboles de decisiones puede manejar datos faltantes sin problemas. A su vez como la cantidad de datos faltantes es muy pequeña y en variables discretas se decidió dejarlos para no interferir en el análisis exploratorio.

Gráfico 3: cantidad de datos faltantes

| | |
|-------------------|---|
| aut_fin | 0 |
| num_emp | 0 |
| usd_emp | 0 |
| cost_pwage | 0 |
| inv_pworks | 0 |
| area | 0 |
| pob_proy | 0 |
| pob_uni (b) | 0 |
| pob_uni | 0 |
| ppr_girs | 0 |
| ton_mes | 0 |
| cap_girs | 0 |
| ord_girs | 0 |
| ord_cobro | 0 |
| agua_potable | 9 |
| serv_barrido | 0 |
| lsb_urbana | 5 |
| lsb_rural | 7 |
| reciclar_ro | 0 |
| cantidad_organico | 0 |
| sitio_deposito | 0 |
| gestion_dsanit | 0 |

Elaboración: Propia

2.3 Descripción de las variables

Tabla 1 variable dependiente

Variable Dependiente

| Código de Variable | Nombre de la Variable | Descripción | Categoría de la Variable |
|--------------------|--|--|---|
| y_pwork | Inversión en obras publicas encima de la media | Variable binaria obtenida a través de la inversión en obras públicas | 1: invierte encima de la media 0: no invierte por encima de la media |

Elaboración: Propia

Tabla 2 variables Independientes

Variables Independientes

| Código de Variable | Nombre de la Variable | Descripción | Categoría de la Variable |
|--------------------|------------------------------|--|--------------------------|
| aut_fin | autonomía financiera | Nivel de autonomía financiera en porcentaje por municipio. | Numérica |
| num_emp | Número de empleados públicos | Número de empleados públicos por municipio | Numérica |
| usd_emp | Dinero por empleado | Cantidad de dinero autofinanciado de los municipios dividido para la cantidad de empleados | Numérica |
| cost_pwage | Gasto en sueldos | Porcentaje de gasto en sueldos en un municipio | Numérica |

| | | | |
|--------------------|---|---|----------|
| area | area | Área en km ² del tamaño del cantón | Numérica |
| pob_proy | Población proyectada | Número total de población proyectada por cantón | Numérica |
| pob_uni | Población de universitarios proyectada | Número total de personas con título universitario por cantón | Numérica |
| ppr_girs | Producción per cápita de residuos | Producción per cápita de residuos por cada municipio | Numérica |
| ton_mes | Toneladas recogidas al mes de residuos | Toneladas el mes recogidas de residuos por municipio. | Numérica |
| lsb_urbana | Longitud de calles susceptibles a barrido urbanas | Longitud de calles susceptibles a barrido dada en km de la zona urbana. | Numérica |
| lsb_rural | Longitud de calles rurales susceptibles a barrido | Longitud de calles susceptibles a barrido dada en km de la zona rural | Numérica |
| cantidad_or | Cantidad de residuos orgánicos | Cantidad de residuos orgánicos recolectadas en el municipio. | Numérica |

Elaboración: Propia

Tabla 3 variables de control

Variables de control

| Código de Variable | Nombre de la Variable | Descripción | Categoría de la Variable |
|---------------------------|---------------------------------|--|---|
| cap_girs | Capacitación manejo de residuos | Variable binaria de si un municipio tiene capacitación de manejo de residuos | 1: cuenta con capacitación 0: no cuenta con capacitación |
| ord_girs | Ordenanza gestión de residuos | Variable binaria de si un municipio cuenta con ordenanza del GIRS | 1: cuenta con ordenanza 0: no cuenta con ordenanza |
| ord_cobro | Ordenanza de cobro | Variable binaria de si un municipio cuenta con ordenanza de cobro. | 1: cuenta con ordenanza de cobro 0: no cuenta con ordenanza de cobro |
| agua_potable | Agua potable | Variable binaria de si un municipio cuenta con agua potable. | 1: cuenta con potable 0: no cuenta con agua potable |
| serv_barrido | Servicio de barrido | Variable binaria que indica si el municipio cuenta con servicio de barrido | 1: cuenta con servicio de barrido 0: no cuenta con servicio de barrido |
| sitio_deposito | Sitio de depósito | Variable Categórica que indica con cuantos | Categórica |

| | | | |
|-----------------------|-----------------------------------|---|---|
| | | sitios de depósito cuenta el municipio. | |
| gestion_dsanit | Gestión de desechos sanitarios | Variable binaria que indica si el municipio cuenta con sistema de gestión de desechos sanitarios. | 1: cuenta con sistema de gestión 0: no cuenta con sistema de gestión |

Elaboración: Propia

2.4 Operacionalización de las variables

2.4.1 Construcción de indicadores y operación de variables

Dentro de las variables obtenidas se tiene la variable de inversión en obras públicas. Antes de conceptualizar esta variable primero es necesario operacionarla:

```
df_cleaned['y_pwork']=(df['inv_pworks']>=df['inv_pworks'].mean())*1
```

La variable resultante de esta operación será una variable binaria en la cual obtendrá el valor de 1 si dicho municipio invierte por encima del promedio en obras públicas y 0 en caso contrario. Esta será nuestra variable de interés y nos permitirá clasificar a los municipios en municipios que invierten por encima de la media o no y analizar las variables que influyen en esta decisión.

También se va a operacionar la variable de tamaño de población proyectada de la siguiente manera:

```
df_cleaned["size_test"]=pd.qcut(df_cleaned.pob_proy,3,[1,2,3])
```

De esta forma se utilizan los cuartiles para obtener 3 grupos con respecto al tamaño de la población donde “1” será pequeño, “2” será mediano y “3” grande. De esta manera podemos separar a los municipios en función de su tamaño poblacional y observar que interacciones se dan en estos niveles. La distribución de los bins estandarizados de la siguiente manera: pequeño=(781.523,9005.38], mediano=(9005.38,22800.617], grande=(22800.617,2700988.057].

Por otro lado, se busca obtener el indicador de densidad poblacional a partir de las variables de población proyectada y área del cantón para obtener esta métrica.

```
df_cleaned["densidad"]=df_cleaned["pob_proy"]/df_cleaned["area"]
```

Se obtiene esta métrica para determinar en la selección de variables si es que podría tomar mayor relevancia que la categorización de municipio solamente por su tamaño poblacional.

Como última variable operacional, también se decidió obtener categorías según los costos en obras públicas para observar de mejor manera las diferencias en los gráficos al dividir los costos según sus cuartiles.

```
df_cleaned["cat_pwage"] =pd.qcut(df_cleaned.cost_pwage,4,[1,2,3,4])
```

De esta manera nos queda que cada categoría corresponde a la siguiente nomenclatura: “1”: bajo: (0.219, 0.41], “2”: medio bajo:(0.41, 0.47], “3”: medio alto: (0.47, 0.55], “4”: alto: (0.55, 0.77].

2.4.2 Conceptualización de variables

Autonomía financiera (aut_fin): nivel de autonomía financiera en porcentaje del municipio.

Numero empleados públicos (num_emp): número de empleados públicos del municipio.

Dinero por empleado (usd_emp): cantidad de dinero autofinanciado de los municipios dividido para cada empleado.

Gasto en sueldos (cost_pwage): porcentaje de cuanto gasta en sueldos un municipio en función de sus gastos totales.

Área: área en km² del tamaño del cantón.

Población proyectada (pob_proy): número total de población proyectada por cantón.

Población de universitarios proyectada (pob_uni): número total de personas con título universitario en adelante por cantón.

Producción per cápita de residuos (ppr_girs): producción per cápita de residuos por cada cantón.

Toneladas recogidas al mes de residuos(ton_mes): toneladas al mes recogidas de residuos por municipio.

Capacitación manejo de residuos (cap_girs): variable categórica donde es 1 si el municipio cuenta con capacitación de manejo de residuos y 0 en caso contrario.

Ordenanza GIRS (ord_girs): variable categórica donde es 1 si el municipio cuenta con ordenanza de GIRS o 0 en caso contrario.

Ordenanza de cobro (ord_cobro): variable categórica donde es 1 si el municipio cuenta con ordenanza de cobro por gestión de residuos o 0 en caso contrario.

Agua potable(agua_potable): variable binaria donde es 1 si el municipio cuenta con agua potable y 0 en caso contrario.

Servicio de barrido(serv_barrido): variable binaria donde es 1 si el municipio cuenta con servicio de barrido y 0 en caso contrario.

Longitud de calles susceptibles a barrido urbanas (lsb_urbana): longitud de calles susceptibles a barrido dada en km de la zona urbana.

Longitud de calles susceptibles a barrido rural (lsb_rural): longitud de calles susceptibles a barrido dada en km de la zona rural.

Recicla residuos orgánicos(reciclar_or): variable binaria donde es 1 si el municipio recicla sus residuos orgánicos y 0 en caso contrario.

Cantidad de residuos orgánicos(cantidad_or): cantidad de residuos orgánicos por municipio.

2.5 Metodología Propuesta

2.5.1 Herramientas y criterios estadísticos

Debido a la naturaleza de los datos limitados que se manejan en el sector público ya sean por falta de estandarización en el manejo de datos a nivel cantonal o falta de reformas en la gestión pública; Se considera la opción de utilizar técnicas que se utilizan en machine learning y pueden ser útiles para realizar estudios econométricos debido a que son robustas para problemas frecuentes como los datos desbalanceados y problemas de outliers. “Most economists are familiar with decision trees that describe a sequence of decisions that results in some outcome. A tree classifier has the same general form, but the decision at the end of the process is a choice about how to classify the observation. The goal is to construct (or “grow”) a decision tree that leads to good out-of-sample predictions.” (Varian, 2014). El problema de esta investigación entra en un problema de clasificación por lo cual es un motivo más para utilizar un árbol de decisiones (CART).

2.5.2 XGBoost para selección de variables

Antes de realizar el clasificador de árbol es importante filtrar las variables según su importancia. Para esta selección de variables se utilizará el método de Extreme Gradient Boosting o xgboost. La biblioteca XGBoost implementa el algoritmo de árbol de decisión de aumento de gradiente. Este algoritmo tiene muchos nombres diferentes, como aumento de gradiente, árboles de regresión aditiva múltiple, aumento de gradiente estocástico o máquinas de aumento de gradiente. El impulso es una técnica de conjunto donde se agregan nuevos modelos para corregir los errores cometidos por los modelos existentes.

Los modelos se agregan secuencialmente hasta que no se puedan realizar más mejoras. Un ejemplo popular es el algoritmo AdaBoost que pondera los puntos de datos que son difíciles de predecir. El aumento de gradiente es un enfoque en el que se crean nuevos modelos que predicen los residuos o errores de modelos anteriores y luego se suman para hacer la predicción final. Se llama aumento de gradiente porque utiliza un algoritmo de descenso de gradiente para minimizar la pérdida al agregar

nuevos modelos. Este enfoque admite problemas de modelado predictivo de clasificación y regresión. A continuación, se presenta la fórmula que utiliza el modelo XGBoost.

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_i \Omega(f_k)$$
$$\text{donde } \Omega(f) = \sqrt{T} + \frac{1}{2} \lambda \|\omega\|^2$$

En la función de pérdida al lado izquierdo se tiene el valor real contrastado con el valor estimado y al lado derecho se tiene una función de regularización donde se toma en consideración el número de modelos finales (T) y $|\omega|$ es el peso de los nodos en cada modelo.

Las medidas para la importancia de variables que se utilizarán serán las de “weight” y “gain”. De esta manera con la escala de weight se podrá revisar las variables con mayor variación en sus datos por lo tanto una mayor probabilidad de explicar la clasificación y en segundo la medida de “gain” permitirá revisar las variables categóricas que debido a su poca variación por tener menor cantidad de valores únicos podrían perderse en la escala anterior.

2.5.3 Índice de diversidad de Shannon

A su vez para medir el balance de los datos, se utilizará el índice de diversidad de Shannon. Generalmente es un índice que se utiliza en estudios ecológicos para ver la diversidad de las especies en un ecosistema. Pero también se podría utilizar para medir el balance de los datos. “Los índices de diversidad son estadísticas que se utilizan para resumir la diversidad de una población en la que cada miembro pertenece a un grupo único. Por ejemplo, en ecología, los grupos son típicamente especies. En ecología, la riqueza de especies se refiere al número de especies y la uniformidad de especies se refiere a la homogeneidad de las especies.” (NIST, 2012)

Para el caso del análisis del balance de los datos, la diversidad de Shannon se aplica de la siguiente manera:

En un data set de n instancias, si se tienen k clases de tamaño c_i entonces:

$$H = - \sum_{i=1}^k \frac{c_i}{n} \log \frac{c_i}{n}$$

Esto resulta en:

0 cuando hay una sola clase, es decir tiende a 0 cuando el data set es muy desbalanceado.

Log k cuando todas las clases son balanceadas en el mismo tamaño n/k .

Por ende, se puede utilizar el siguiente estimador para medir el balance:

$$Balance = \frac{H}{\log k} = \frac{- \sum_{i=1}^k \frac{c_i}{n} \log \frac{c_i}{n}}{\log k}$$

Y esto resultaría en:

0 para una base de datos desbalanceada

1 para una base de datos balanceada

2.5.4 Árbol de Clasificación y Regresión (CART)

Una vez identificadas las variables según su importancia se puede proceder a la construcción del árbol. El árbol de decisión es un modelo no paramétrico que se utiliza para realizar predicciones mediante observaciones de un objeto y culminar en conclusiones del mismo objeto. En los casos en que la variable a estudiar tenga valores discretos, se pasa a llamar un modelo de árbol clasificador. Según el journal de Data Mining Classification: Basic Concepts, Decision Trees, and Model Evaluation por Steinbach (2004). Se reconocen los siguientes términos en un árbol de decisiones:

Root node: representa toda la población o muestra y esta se divide en dos o más conjuntos homogéneos.

Splitting: es un proceso de división de un nodo en dos o más subnodos.

Decision node: cuando un subnodo se divide en otros subnodos, se denomina nodo de decisión.

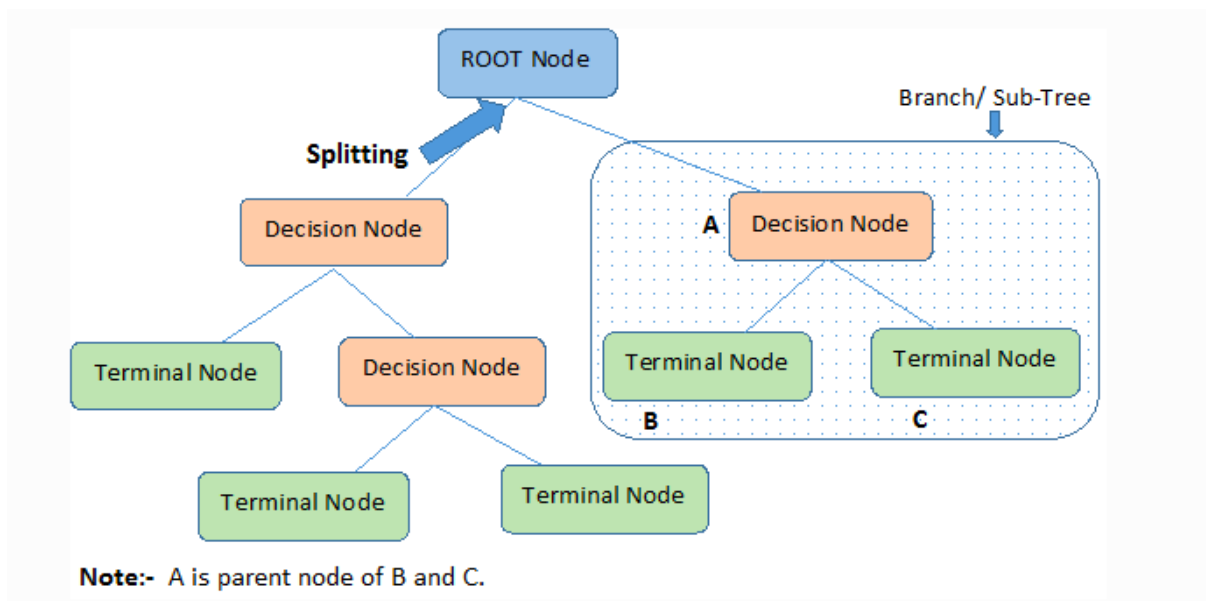
Leaf/terminal: los nodos que no se dividen se denominan nodo hoja o terminal.

Pruning: cuando eliminamos subnodos de un nodo de decisión, este proceso se llama poda. Se puede decir el proceso opuesto a la división.

Branch/subtree: una subsección de todo el árbol se denomina rama o subárbol.

Parent and Child Node: un nodo, que se divide en subnodos, se denomina nodo padre de subnodos, mientras que los subnodos son hijos de un nodo padre.

Gráfico 4 Estructura de árbol



Elaboración: obtenida de artículo online Decision Tree Algorithm (Chauhan, 2020)

El árbol de decisiones como clasificador (CART) no utiliza la varianza como criterio para clasificar la data, por lo que no es sensible a la presencia de outliers. En esta investigación se utiliza el criterio de Gini para realizar las clasificaciones de la data.

2.5.5 Índice de GINI

El índice de Gini mide el índice de pureza de un nodo del árbol. El cual mide la probabilidad de no sacar dos registros de la misma clase del nodo. De esta manera se podría decir que a mayor índice de Gini menor pureza, por lo que se irán seleccionando las variables con un menor índice de Gini. El índice de Gini está representado de la siguiente manera:

$$GINI(t) = 1 - \sum_{i=1}^n (P_i)^2$$

Donde P_i es la probabilidad de que un ejemplo sea de la clase i .

Una vez obtenido el árbol de clasificación, se procede a utilizar métricas para medir su precisión y desempeño. Una de las mejores métricas para medir el desempeño de un árbol de decisiones vendría a ser la matriz de confusión.

2.5.6 Matriz de confusión

En la investigación de Josip Mesaric y Dario Sebalj, se utilizó la matriz de confusión para medir el desempeño de los diferentes modelos de árboles que utilizaron para predecir el éxito académico de los estudiantes en la universidad de Osijek. (Josip Mesari, 2016)

La matriz de confusión nos muestra cuatro valores: verdadero positivo(tp), falso positivo(fp), falso negativo(fn) y verdadero negativo(tn).

Gráfico 5 Matriz de confusión

| | | Clase Verdadera | |
|-----------------|----------|-----------------|----------|
| | | Positivo | Negativo |
| Clase Predecida | Positivo | VP | FP |
| | Negativo | FN | VN |

Elaboración: obtenida de artículo online: Confusion Matrix (Mohajon, 2020)

Gráfico 6 Medidas Matriz de Confusión

| | | |
|---|--|--|
| Precision = $\frac{\text{Verdadero positivo}}{\text{Resultado actual}}$ | $\frac{\text{Verdadero positivo}}{\text{VP+FP}}$ | F1 = $\frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$ |
| Recall = $\frac{\text{Verdadero positivo}}{\text{Resultado predecido}}$ | $\frac{\text{Verdadero positivo}}{\text{VP+FN}}$ | |
| Accuracy = | $\frac{\text{VP+VN}}{\text{TOTAL}}$ | |
| Especificity = | $\frac{\text{Verdadero negativo}}{\text{VN+FP}}$ | |

Elaboración: obtenida de artículo online: Confusion Matrix (Mohajon, 2020)

En el caso de esta investigación los verdaderos positivos vendrían a ser los municipios que si invierten por encima de la media y los verdaderos negativos son los municipios que no invierten por encima de la media. Por último, se utilizará la medida de accuracy y el F1 score para comparar los distintos árboles.

CAPITULO III

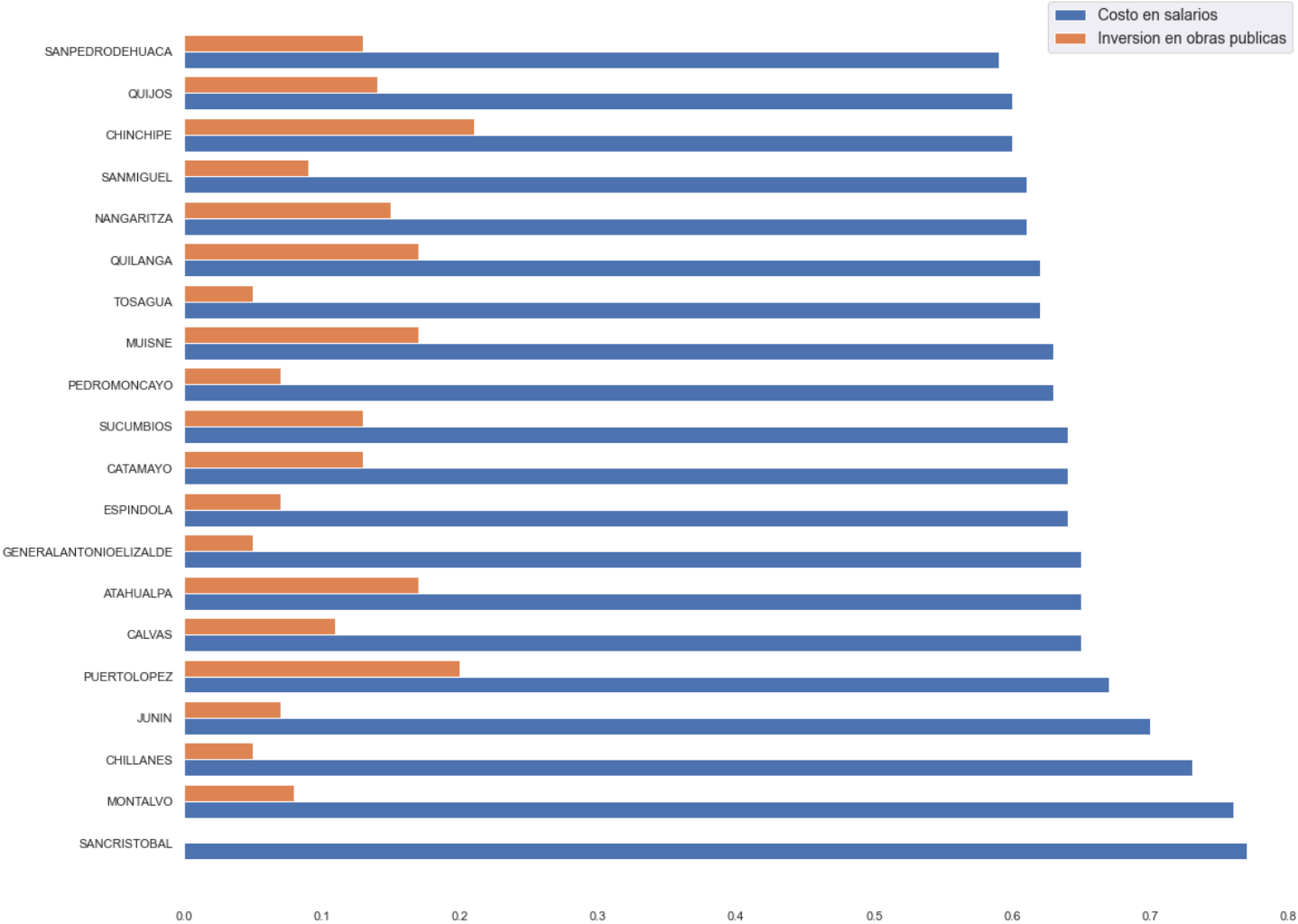
3. RESULTADOS

3.1.1 Análisis Exploratorio

Al realizar el análisis exploratorio de los datos se encontró que los municipios que cuentan con la mayor cantidad de gasto en salarios también tenían un bajo porcentaje de inversión en obras públicas.

Gráfico 7 top 20 municipios en costo de salarios

20 primeros municipios con mayor costo en salarios



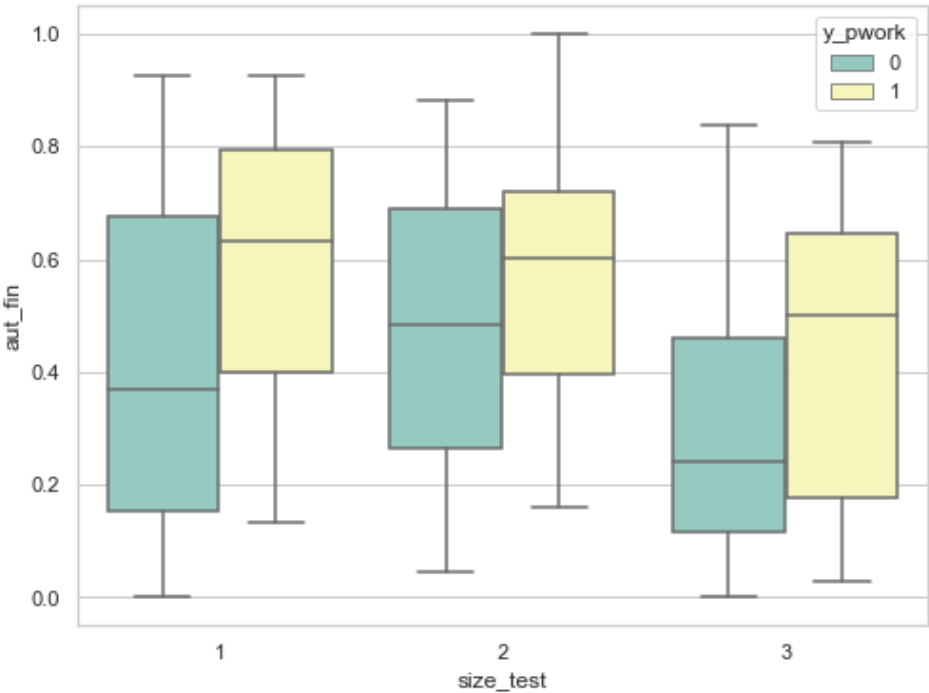
Elaboración: Propia

Incluso se encontró que en el municipio de San Cristóbal se ha invertido un 0% en obras públicas a pesar de que sea el municipio con el mayor gasto en salarios superando la marca de 75% de gastos enfocados en este rubro.

3.1.2 Distribución de variables categóricas

A continuación, se revisará la distribución de los datos y detección de outliers. En el primer gráfico se realiza en función del tamaño del municipio y su nivel de autonomía financiera, también se diferencia con diferentes colores a la distribución de los municipios que invierten por encima de la media de color amarillo y los que no invierten por encima de la media están de color verde. En la siguiente figura la variable size_test divide en 3 cuartiles los tamaños de los cantones.

Gráfico 8 Boxplot 1



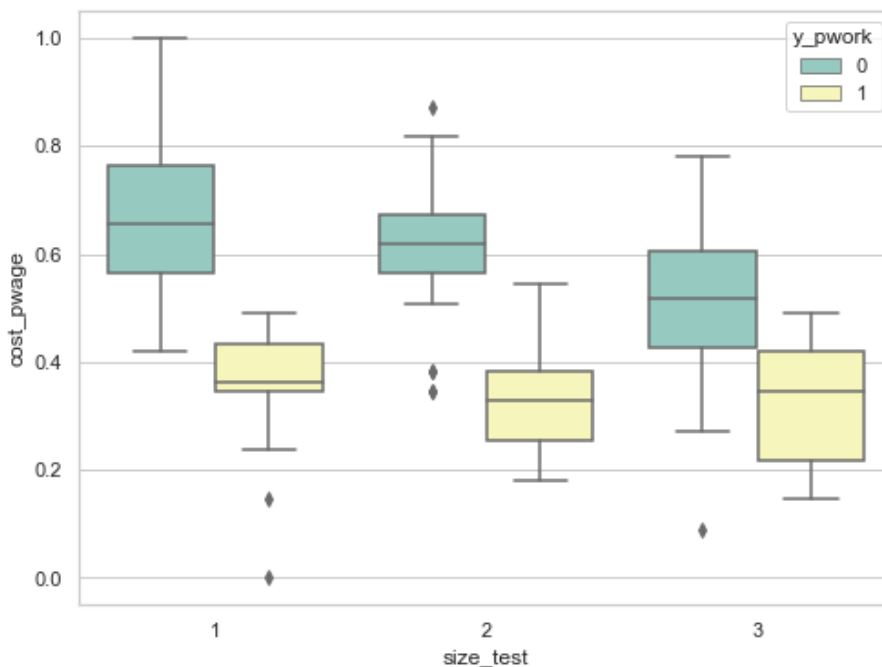
Elaboración: propia

A simple vista se puede observar que no existen outliers en estas variables. También se puede decir que los municipios que si invierten por encima del promedio

en obras públicas tienen su media de la autonomía financiera por encima de los municipios que no invierten más sin importar el tamaño del municipio.

A continuación, se realiza un gráfico similar en donde se buscará observar la distribución del costo en salarios en función del tamaño del municipio, diferenciando los municipios que invierte más en obras públicas y los que no. A pesar de que en este gráfico se encuentre la presencia de outliers, no es necesario tomar medidas para eliminarlos debido a que el modelo de clasificador de árbol es robusto a los outliers y es mucho mejor tener una mayor cantidad de datos.

Gráfico 9 Boxplot 2



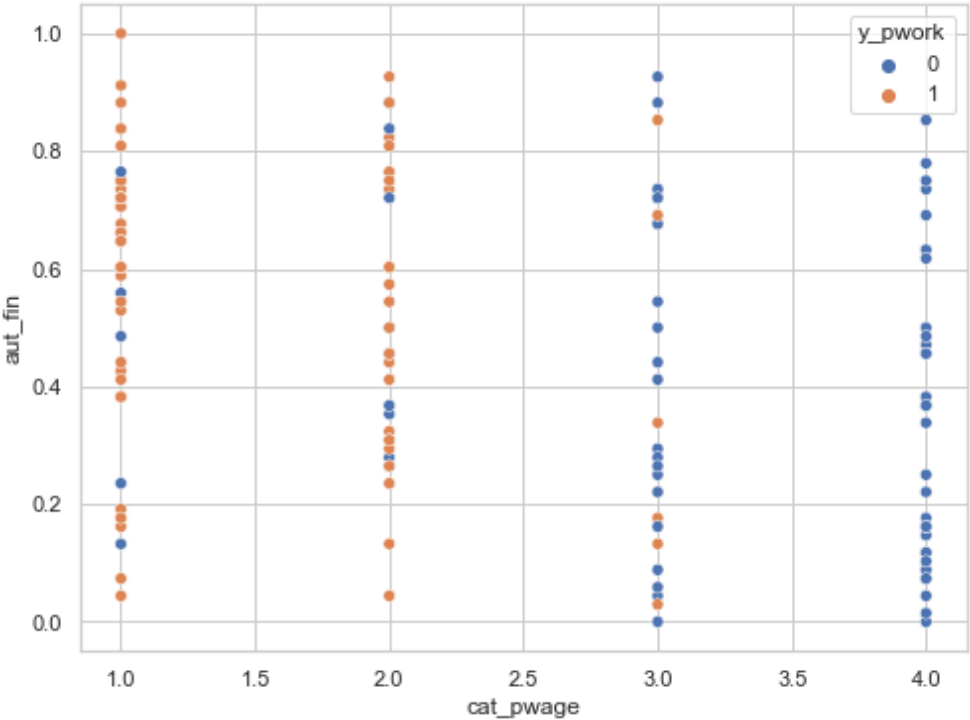
Elaboración: propia

Luego de revisar como están distribuidos los datos en función de sus cuartiles, se procede a revisar la relación entre las variables utilizando scatterplots.

En este primer grafico se observará la relación que existe entre el costo de salarios y la autonomía financiera separando los grupos de datos en función de la variable dependiente.

A simple vista se puede observar que la autonomía financiera no tiene mayor diferencia en su distribución al compararla con los costos de salarios. Pero si se puede observar una clara distribución de los municipios que invierten más en obras públicas que se concentran en las categorías donde el costo de salarios es menor.

Gráfico 9 scatter plot 1

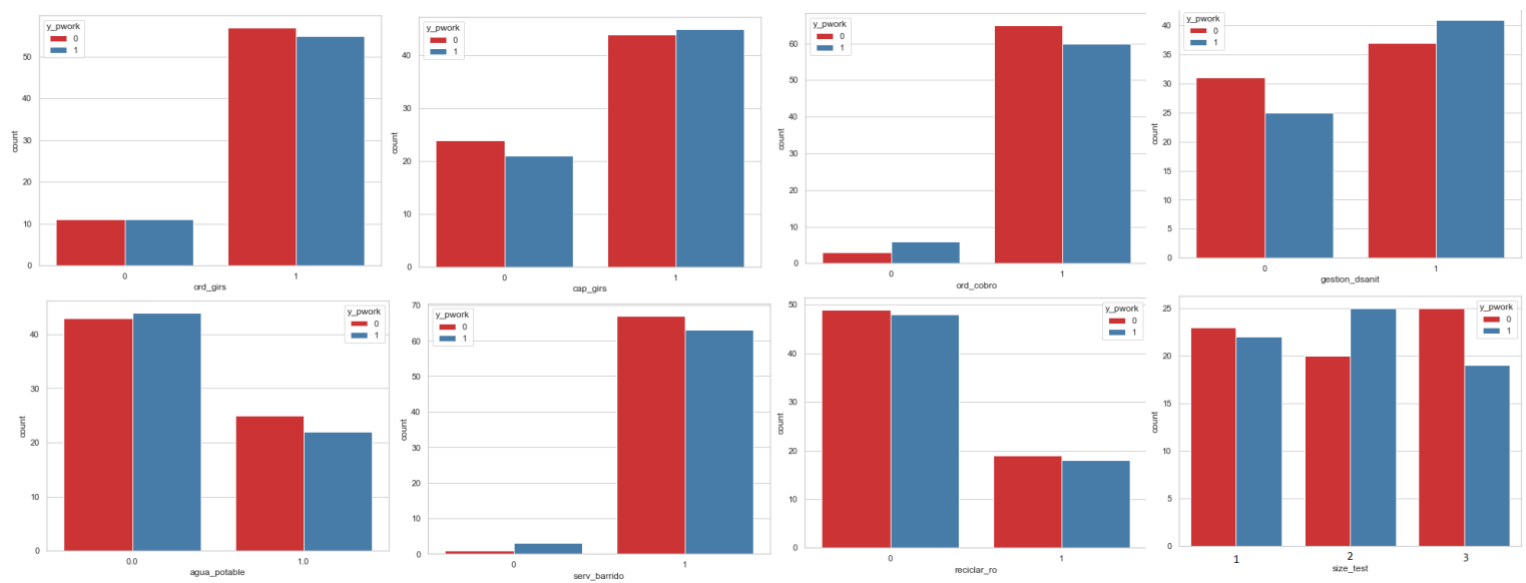


Elaboración: propia

3.1.3 Revisión de asimetría de variables categóricas

En esta sección se realizarán gráficos de barras para revisar la asimetría de las variables categóricas en función de los dos grupos de municipios según la variable dependiente.

Gráfico 10 simetría de variables



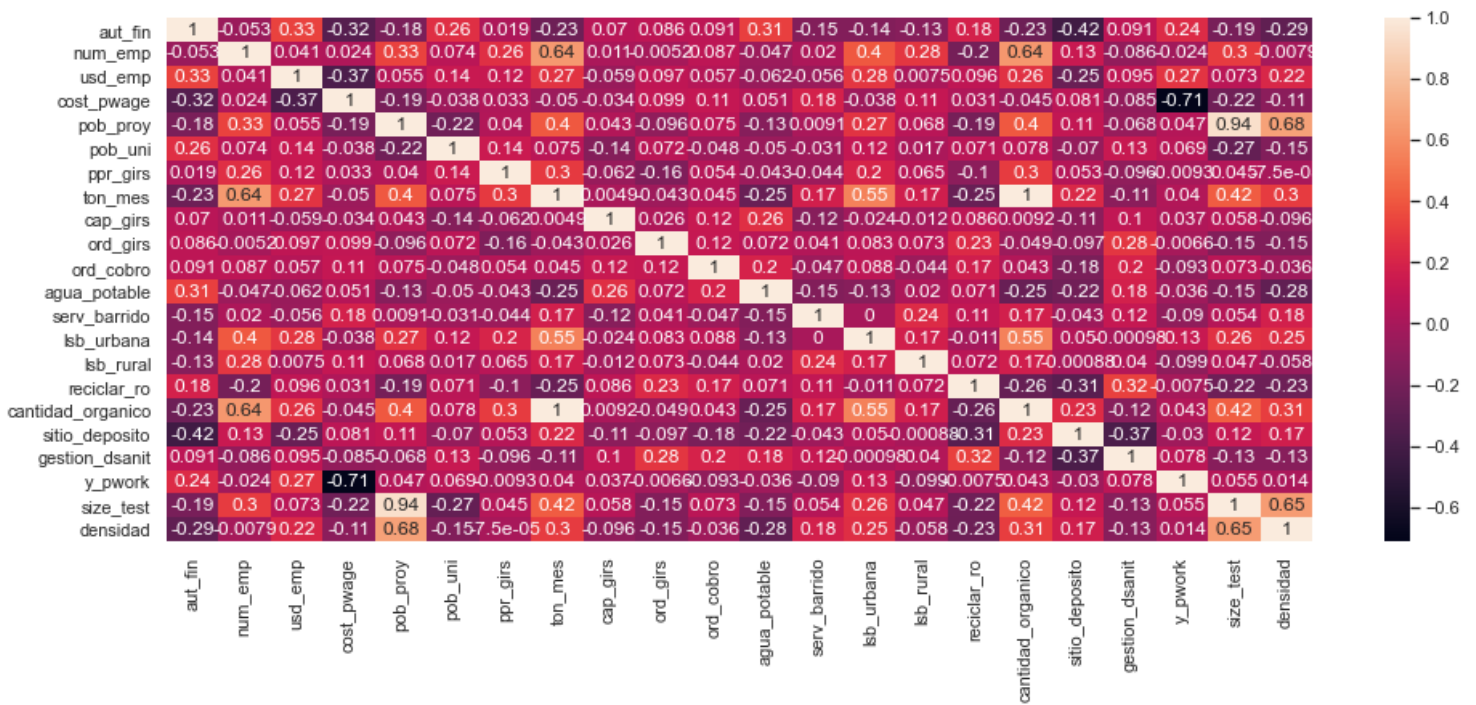
Elaboración: propia

En los gráficos se puede observar que algunas de las variables categóricas como las de servicio de barrido, ordenanzas de cobro y ordenanza de gestión de residuos son asimétricas. A pesar de que estas variables tengan pocos datos correspondientes a uno de los grupos de la variable dependiente, el modelo clasificador es robusto a las asimetrías por lo que es bastante probable que dichas variables que no aporten de manera significativa al modelo, no se terminen considerando al final.

3.1.4 Matriz de correlaciones

También se realiza una matriz de correlaciones para detectar que variables podrían ser redundantes y no estarían agregando nada al modelo del clasificador de árbol.

Gráfico 11 Matriz de Correlaciones 1



Elaboración: propia

Las variables que se decidieron eliminar debido a su redundancia son: "pob_uni(b)", "inv_pworks", "cat_pwage", "size_dens", "area", "lsb_urbana", "lsb_rural". Se eliminaron estas variables debido a que existen otras que ya explican la relación con la variable dependiente y también porque algunas de estas sirvieron para la construcción de variables categóricas como las del tamaño de municipio y costos en salarios.

3.2 Selección de variables (XGBoost)

Para la selección de variables que se utilizarán en el modelo de clasificador de árbol se utilizara el método de XG boost. Primero se corre una función para determinar el número de variables que se deberían utilizar para obtener la mejor precisión en el modelo.

Tabla 4 XGBoost número de variables

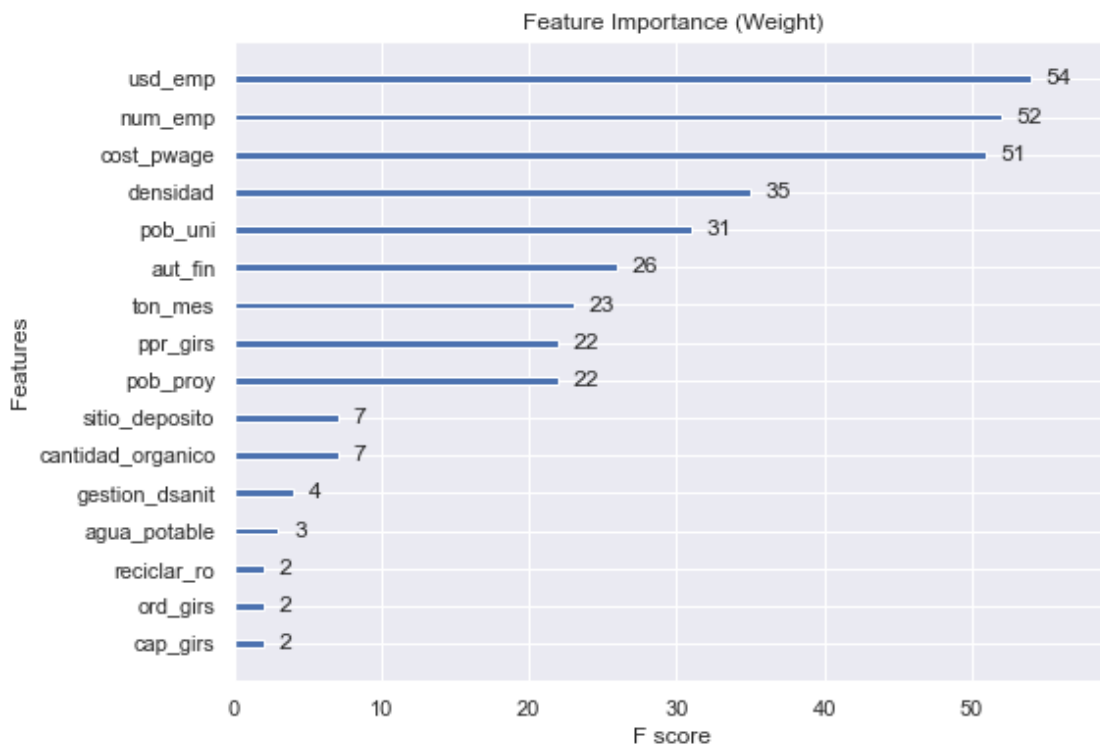
| | | |
|---------------------|-------------|-------------------------|
| Thresh=0.000 | n=19 | Accuracy: 84.44% |
| Thresh=0.006 | n=13 | Accuracy: 84.44% |
| Thresh=0.006 | n=12 | Accuracy: 86.67% |
| Thresh=0.019 | n=11 | Accuracy: 88.89% |
| Thresh=0.020 | n=10 | Accuracy: 86.67% |
| Thresh=0.033 | n=9 | Accuracy: 86.67% |
| Thresh=0.034 | n=8 | Accuracy: 80.00% |
| Thresh=0.051 | n=7 | Accuracy: 77.78% |
| Thresh=0.054 | n=6 | Accuracy: 84.44% |
| Thresh=0.071 | n=5 | Accuracy: 82.22% |
| Thresh=0.085 | n=4 | Accuracy: 80.00% |
| Thresh=0.104 | n=3 | Accuracy: 75.56% |
| Thresh=0.130 | n=2 | Accuracy: 75.56% |
| Thresh=0.386 | n=1 | Accuracy: 88.89% |

Elaboración: Propia

En esta tabla se observa que desde 12 variables hasta 9 se tiene una buena precisión del modelo. En este caso, siendo mejor el modelo que utilice 11 variables puesto que daría una precisión de 88.89%.

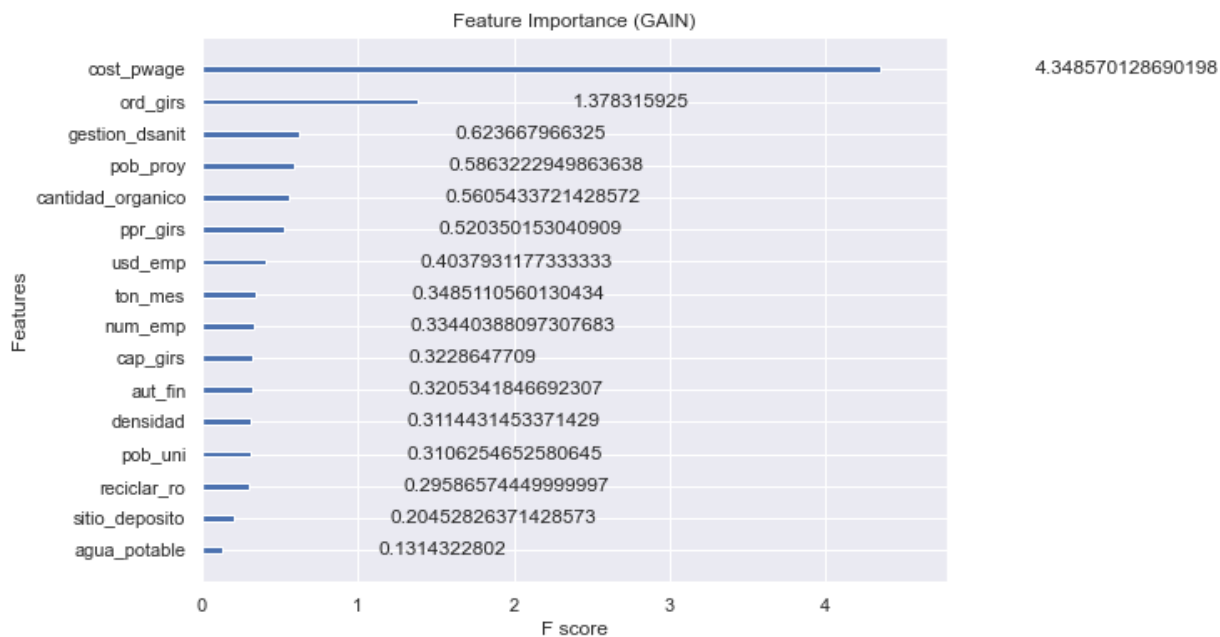
Para obtener las variables que se van a utilizar se utilizaran otras métricas para medir su aporte al modelo. Dentro de las métricas a tomar en cuenta para evaluar las variables de importancia se utiliza "weight" y "gain". Weight le da mayor importancia a las variables que tienen mayor varianza consigo mismas, es decir la cantidad de información que se pueda obtener de una variable. El problema con esta medida es que no toma en cuenta a las variables categóricas debido a su poca cantidad de valores únicos.

Gráfico 12 Importancia de variables(weight)



Elaboración: propia

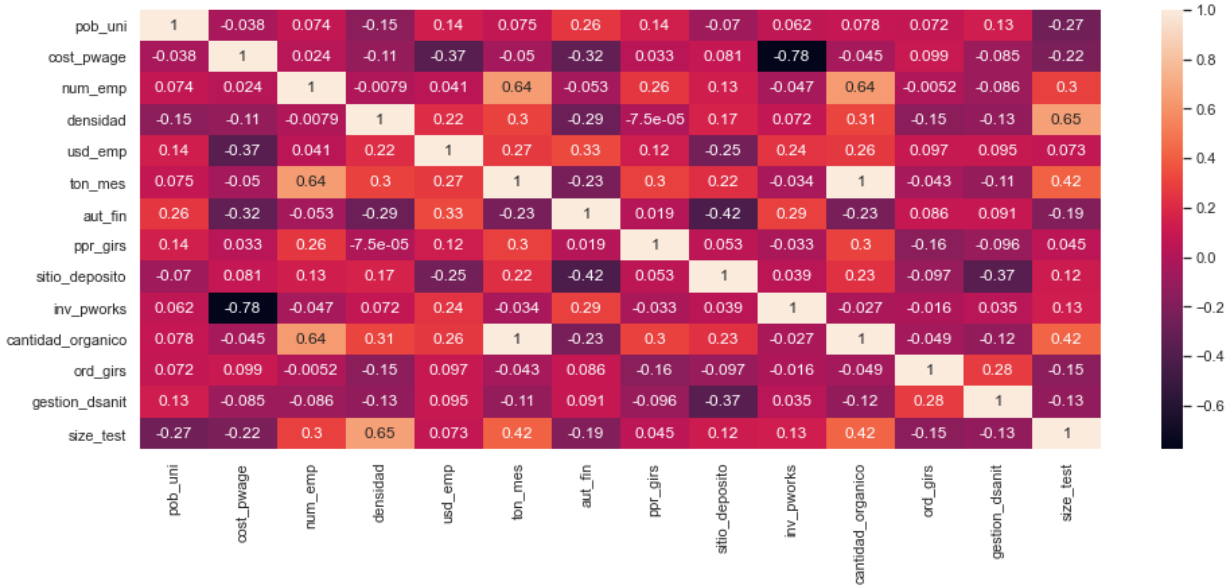
Gráfico 12 Importancia de variables (Gain)



Elaboración: propia

Para obtener la importancia de las variables categóricas se utiliza el método de “gain”. De esta forma se puede ordenar las variables en función de cuanto aportan a la predicción de la variable dependiente. Después de observar los dos gráficos de “gain” y “weight”. Se decide seleccionar las siguientes variables. Después de filtrar las variables redundantes y obtener las variables con mayor importancia se volvió a realizar el mapa de correlaciones.

Gráfico 13 Matriz de Correlaciones 2



Elaboración: propia

Se puede destacar que la variable que mayor correlación tienen con la inversión en obras públicas es la de los costos de salarios y tiene una correlación negativa. De ahí le siguen con correlaciones positivas la autonomía financiera y el dinero autofinanciado por empleado, así como la producción per cápita de residuos y el sitio de depósito. Se puede observar que la mayoría de las variables mantienen una correlación alta con la variable objetivo. Utilizando el modelo de árbol clasificador se podrá profundizar exactamente como estas variables y en qué puntos afectan a la toma de decisiones al momento de invertir en obras públicas.

Tabla 5 variables seleccionadas

| Código de Variable | Nombre de la Variable | Descripción | Categoría de la Variable |
|---------------------------|--|--|---------------------------------|
| aut_fin | autonomía financiera | Nivel de autonomía financiera en porcentaje por municipio. | Numérica |
| num_emp | Número de empleados públicos | Número de empleados públicos por municipio | Numérica |
| usd_emp | Dinero por empleado | Cantidad de dinero autofinanciado de los municipios dividido para la cantidad de empleados | Numérica |
| cost_pwage | Gasto en sueldos | Porcentaje de gasto en sueldos en un municipio | Numérica |
| pob_uni | Población de universitarios proyectada | Número total de personas con título universitario por cantón | Numérica |
| ppr_girs | Producción per cápita de residuos | Producción per cápita de residuos por cada municipio | Numérica |
| ton_mes | Toneladas recogidas al mes de residuos | Toneladas el mes recogidas de residuos por municipio. | Numérica |
| cantidad_organico | Cantidad de residuos orgánicos | Cantidad de residuos orgánicos | Numérica |

| | | | |
|-----------------------|--------------------------------|---|---|
| | | recolectadas en el municipio. | |
| ord_girs | Ordenanza gestión de residuos | Variable binaria de si un municipio cuenta con ordenanza del GIRS | 1: cuenta con ordenanza 0: no cuenta con ordenanza |
| sitio_deposito | Sitio de depósito | Variable Categórica que indica con cuantos sitios de depósito cuenta el municipio. | Categórica |
| gestion_dsanit | Gestión de desechos sanitarios | Variable binaria que indica si el municipio cuenta con sistema de gestión de desechos sanitarios. | 1: cuenta con sistema de gestión 0: no cuenta con sistema de gestión |

Elaboración: Propia

3.3 Árbol de Clasificación

Para medir el desempeño de los diferentes modelos se utilizarán las medidas de clasificación correcta. En la siguiente tabla se muestran los resultados de los modelos de árbol de clasificación.

Tabla 6 Comparación de modelos

| Decision Tree Model | number of leaves | size of tree | Correctly Classified (Accuracy) | Incorrectly Classified |
|------------------------------|-------------------------|---------------------|--|-------------------------------|
| j48 | 2 | 3 | 118 (88.05%) | 16 (11.94%) |
| CART (W/ cost_pwage) | 5 | 9 | 123(91.79%) | 11(8.21%) |
| CART (W/o cost_pwage) | 8 | 15 | 105(78.35%) | 29(21.64%) |

Elaboración: Propia

En la tabla anterior se exponen los resultados de dos modelos de árbol, uno que utiliza un modelo de j48 y otro que utiliza un modelo CART, También existe uno extra el cual es una variación del primer modelo de CART al no incluir la variable con mayor correlación con la variable dependiente.

Se decide no utilizar el modelo j48 debido a que, a pesar de su alto porcentaje de clasificación correcta, el árbol resultante solo contiene un nodo raíz sin nodos hijos. Esto implicaría que no se podría obtener una interpretación que permita explicar exactamente la interacción entre variables que contribuyen a la inversión en obras públicas por parte de los municipios.

Debido a la métrica de accuracy no es la mejor forma de comparar el desempeño de distintos modelos de clasificador de árbol, se utilizará el F1 score debido a que es robusto a la asimetría de datos. A continuación, se presenta la tabla de las distintas métricas para los modelos en cuestión incluyendo el F1 score:

Tabla 7 resumen mejores modelos

| Decision Tree Model | accuracy | bal_accuracy | sensitivity | specificity | precision | F1 Score |
|-----------------------------|-----------|--------------|-------------|-------------|-----------|----------|
| CART(W/cost_pwage) | 0.9179104 | 0.9184492 | 0.8823529 | 0.9545455 | 0.952381 | 0.916031 |
| CART(W/o cost_pwage) | 0.7910448 | 0.7909982 | 0.7941176 | 0.7878788 | 0.794118 | 0.794118 |

Elaboración: Propia

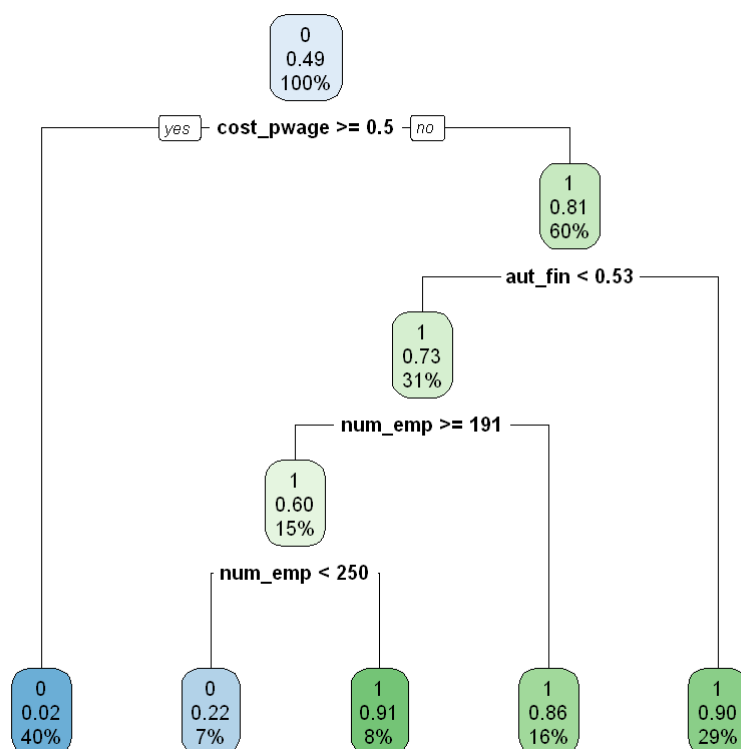
Al comparar el F1 score con la accuracy del modelo se puede observar que nuestro modelo ha sido relativamente robusto a la asimetría de los datos en ciertas variables categóricas. Se obtiene de igual manera que el mejor modelo en función de su precisión y correcta clasificación sigue siendo el primer modelo CART que incluye el costo de salarios en sus variables independientes.

3.3.1 Interpretación de los árboles

Para la interpretación de los árboles se utilizarán los que brindaron los mejores resultados tanto en capacidad de explicación, así como su puntaje de predicción.

El primer árbol que se interpretara es el árbol CART que incluye la variable de costos de salarios.

Gráfico 14 *Árbol de decisiones CART(w/costo de salarios)*



Elaboración: propia

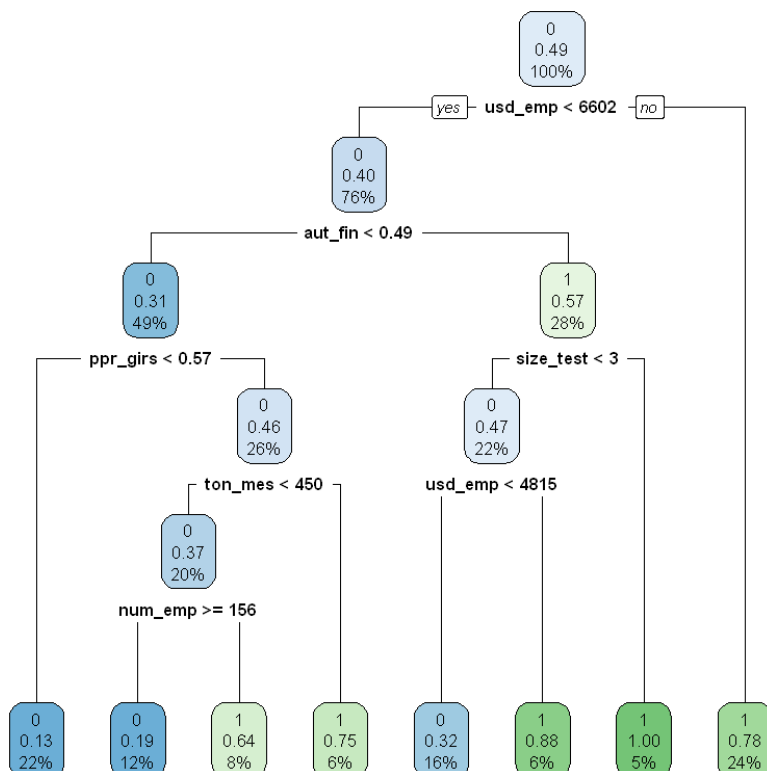
Se puede observar una historia bastante marcada al examinar como están distribuidas las hojas del árbol y sus puntos de quiebre. Primero se observa que si un municipio tiene gastos en salarios equivalentes o superiores al 50% de sus gastos entonces el municipio no invertirá más en obras públicas, para mayor significancia a este punto también se puede agregar que el 40% de la muestra se concentra en ese resultado. Luego en caso contrario de que un municipio tenga un menor porcentaje de gastos en salarios, entonces se procede a revisar la autonomía financiera del mismo. En caso de que la autonomía financiera sea mayor al 53% de los ingresos de dicho municipio entonces invierte por encima de la media en obras públicas; en este resultado se encuentra la segunda mayor distribución de los datos de la muestra. En caso de que no cumpla con ninguna de las condiciones mencionadas antes, entonces se procede a revisar el número de empleados lo cual se podría resumir de que si un municipio no excede en 191 empleados entonces si invierte y en el caso de los exceda

o sean igual a ese número entonces no invertirá por encima de la media en obras públicas.

De manera resumida se puede observar que la inversión del municipio depende mucho de sus gastos corrientes. Es decir, si el monto financiero de los municipios se ve afectado por los gastos en la masa salarial entonces es muy difícil que el municipio vaya a invertir más en obras públicas debido a que no tendría de dónde invertir. A su vez, podemos ver que la autonomía financiera ayuda a mitigar un poco este problema, pero como se puede observar si el costo de salarios excede o equipará al 50% de los gastos en el municipio entonces sin importar el nivel de autonomía financiera, el municipio no invertirá más en obras públicas.

A continuación, se revisarán los resultados del segundo árbol con menor puntaje de predicción debido a que no se incluyó el costo de salarios, pero permite observar la interacción entre otras variables.

Gráfico 15 *Árbol de Decisiones(W/o costo de salarios)*



Elaboración: propia

En este árbol de decisiones se mantienen la mayoría de los principios del árbol anterior, pero ofrece una interpretación extra debido a la inclusión de otras variables. Primero se puede observar que en este árbol el nodo base es el colchón financiero del cual dispone un municipio, es decir la cantidad de dinero que pueden autogenerar en función de sus empleados. Si el municipio puede autogenerar más de 6602 dólares por empleado entonces este municipio si invierte por encima de la media en obras públicas. Por otro lado, el otro resultado en donde se encuentra la segunda concentración más grande de los datos es el cual resulta de la incapacidad de obtener el colchón financiero suficiente y una autonomía financiera adecuada, lo cual resulta en revisar el nivel de producción per cápita de residuos en el cantón. En caso de que esta producción per cápita de residuos sea menor a 0.57 entonces no se invierte por encima de la media en obras públicas.

La interpretación de este árbol resulta ser un poco más complicada que el anterior debido a que se comparan variables ambientales con variables de gastos corrientes. Pero si se profundiza un poco más, se puede llegar a la intuición de que, si las obligaciones que tienen un municipio con la comunidad no son tan grandes, es decir si la cantidad de residuos de los que se tiene que hacer cargo el municipio no excede cierto punto entonces por lógica invierte menos en dicho rubro por lo cual vendría a ser menos inversión en obras públicas relativamente al de los gastos salariales. A su vez, se puede observar que en este árbol también aparece la significancia del tamaño del municipio diferenciando a los municipios más grandes de los pequeños y como si un municipio es más grande tiende a invertir más en obras públicas al cumplirse algunos requisitos.

CAPITULO IV

4. CONCLUSIONES Y RECOMENDACIONES

Al observar los resultados obtenidos de los árboles de decisiones, se puede inferir que la mayoría de los municipios se ven limitados en la capacidad de reinvertir en obras publicas debido a sus altos niveles de gasto corriente, en donde el gasto de la masa salarial puede llegar a alcanzar el 50% de los gastos de dichos municipios. En caso de que los municipios no sean capaces de obtener un colchón financiero de dinero autofinanciado en función de la cantidad de empleados que tienen, entonces las probabilidades de que estos municipios logren invertir más en obras públicas son muy bajas inclusive llegando a ser nula.

Otro aspecto importante que afecta en la toma de decisiones de los municipios es la cantidad de empleados que tienen, puesto que si exceden los 191 el municipio tiende a invertir menos en obras públicas. La intuición detrás de esto viene del problema de una ineficiencia gerencial puesto que el financiamiento de los municipios en su mayoría viene del gobierno central entonces no existen incentivos a que se busque un punto de equilibrio ente la cantidad de empleados que realmente se necesitan para maximizar el aspecto operacional de dicho municipio. A diferencia de una empresa privada en donde ellos cubren con el financiamiento de su parte operativa y esto les incentiva a buscar la cantidad de empleados suficientes para poder cumplir con su rubro de forma eficiente.

Por último, se tiene un aspecto interesante que se encontró en los árboles, el cual es la cantidad de residuos orgánicos y como cuando un cantón genera más desechos entonces el municipio se ve inclinado a invertir más en obras publicas debido a que invierten más recursos en el manejo de dichos residuos. Aunque no se puedan dar aseveraciones totalmente certeras respecto a este caso debido a que si se incluyeran otro tipo de aspectos como la salud y ámbitos de construcción publica, es posible que la parte ambiental pierda relevancia en función de esos otros datos.

A pesar de esto, se puede decir que la intuición detrás de estos resultados presentes radica en que una de las necesidades más básicas que tiene un municipio por cubrir es el manejo de residuos puesto que es una necesidad inmediata de la cual se debe manejar cada día.

Al comparar estos resultados con investigaciones cuantitativas pasadas se concluye que la capacitación técnica y las habilidades de gerencia de un municipio efectivamente influyen mucho en su propensión a la reinversión.

Así como en estudios pasados se recomendaría que en las encuestas que se realizan por parte del gobierno central a los municipios se incluyan preguntas que ayuden a medir el nivel de necesidad por capacidades técnicas en distintas áreas de los municipios. Por otro lado, un problema que se encontró al realizar la investigación es la falta de información debido a que los municipios no se manejan con un procedimiento estandarizado para el almacenamiento de sus datos y los datos que están disponibles al público no son suficientes para poder realizar investigaciones más profundas.

De igual manera, a pesar de estos resultados encontrados, cabe recalcar que este estudio es solo el inicio de muchos que se deberían realizar para obtener una explicación de qué es lo que realmente sucede con el manejo de los recursos en los municipios y las razones por las cuales, en la mayoría de los casos, la atención en las necesidades de los habitantes es ineficiente.

Este trabajo de investigación busca promover a la actualización de los procedimientos y transparencia de datos para elevar la situación del país a una donde se puedan tomar decisiones basadas en datos y en la cual sea posible combatir la corrupción cuya practica se promueve al tener datos difíciles de contrastar y con cuentas poco transparentes.

5. BIBLIOGRAFÍA

- Alberto Porto, M. G. (2018). Performance and efficiency measures in the public sector. The case of Bolivia. *RAE*, 120-122.
- Alberto Porto, M. G. (2018). Performance and efficiency measures in the public sector. The case of Bolivia. *RAE*, 133-136.
- Casas, A. Y. (2008). LA CAPACIDAD Y DESEMPEÑO DE LOS GOBIERNOS LOCALES: EVIDENCIA DE LAS MUNICIPALIDADES PERUANAS. *CAF*, 20-25.
- Chauhan, N. S. (2020). *kdnuggets*. Obtenido de *kdnuggets*:
<https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html>
- COOTAD. (2010). CODIGO ORGANICO DE ORGANIZACION TERRITORIAL.
- ECUADOR, C. R. (2008). Obtenido de <https://educacion.gob.ec/wp-content/uploads/downloads/2012/08/Constitucion.pdf>
- ECUADOR, C. R. (2008). Obtenido de <https://educacion.gob.ec/wp-content/uploads/downloads/2012/08/Constitucion.pdf>
- Frankle, P. &. (2007). Análisis de la eficiencia del gasto municipal y de sus determinantes. *RePEc*, 148-149.
- Gad, S. d. (2015). Obtenido de <https://www.planificacion.gob.ec/wp-content/uploads/2015/09/Nuevo-Sistema-de-Transferencias-a-Favor-de-los-GAD.pdf>
- Guerra, S. (14 de August de 2020). Obtenido de Primicias:
<https://www.primicias.ec/noticias/economia/municipios-pequenos-gasto-ineficiente-recursos/>
- Guerrero, O. (2007). *omarguerrero*. Obtenido de
<http://www.omarguerrero.org/libros/papp200.pdf>
- Josip Mesari, D. Š. (2016). Decision trees for predicting the academic success of . *Croatian Operational Research Review* , 368-345.
- Margarita Panchana P., H. C. (2017). La gestión de los ingresos y su incidencia en la inversión pública del Gobierno Autónomo Descentralizado Municipal de Salinas. *Revista Científica y Tecnológica UPSE*, 78-83.
- Mohajon, J. (28 de Mayo de 2020). *Towards data science*. Obtenido de Towards data science: <https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826>
- NIST. (2012). *NIST*. Obtenido de National institute of standards and technology:
<https://www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/shannon.htm>
- Plan Nacional de Descentralización. (2013-2015). Obtenido de
<https://www.planificacion.gob.ec/wp-content/uploads/downloads/2012/08/Plan-Nacional-de-Descentralizaci%c3%b3n-2012-2015.pdf>

- SENPLADES. (2012). *planificacion gob.* Obtenido de <https://www.planificacion.gob.ec/wp-content/uploads/downloads/2016/03/GUIA-DE-PLANIFICACION-INSTITUCIONAL.pdf>
- Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *ournal of Economic Perspectives—Volume 2*, 3-28.
- Cpccs.gob.ec. 2010. CODIGO ORGANICO DE ORGANIZACION TERRITORIAL, COOTAD. [online] Available at: <<https://www.cpccs.gob.ec/wp-content/uploads/2020/01/cootad.pdf>> [Accessed 7 September 2021].
- Oas.org. 2021. CONSTITUCION DE LA REPUBLICA DEL ECUADOR 2008. [online] Available at: <https://www.oas.org/juridico/pdfs/mesicic4_ecu_const.pdf> [Accessed 7 September 2021].