Solución del problema de asignación de turnos de personal a través de aprendizaje reforzado Q-Learning. Estudio de caso estación de servicio automotriz Transpiedecuesta S.A.[1]

Solution of the problem of assigning staff shifts through Q-Learning reinforcement learning. Case study of automotive service station Transpiedecuesta S.A.

J. D Mendivelso Velandia[1], M. M. Mora Carvajal[2], N. R. Ortíz Pimiento[3]

[1] Universidad Industrial de Santander, Colombia Email: juan.mendivelso@correo.uis.edu.co
[2] Universidad Industrial de Santander, Colombia Email: michael.mora@correo.uis.edu.co
[3] Universidad Industrial de Santander, Colombia Email: nortiz@uis.edu.co

RESUMEN

Esta investigación aborda el problema de asignación de turnos de personal en una estación de servicio (EDS) llamada Transpiedecuesta SA. La solución propuesta consta de dos fases. En la primera fase, se utiliza un modelo de aprendizaje automático para modelar la demanda de combustible, pronosticando así los requisitos de personal hora a hora. La segunda fase implica la implementación de un algoritmo de aprendizaje reforzado Q-Learning en Python, entrenando a un agente con una función de recompensa y políticas para asignar eficientemente empleados en diferentes turnos durante 3 meses. Los resultados muestran un cronograma de asignación sin turnos fijos, con métricas adecuadas para aproximadamente 16 empleados y un costo total de asignación menor al promedio de la nómina real de la empresa. En términos de costo computacional, se emplearon 25 minutos y un máximo de 1,5 GB de memoria RAM para realizar 1200 episodios de simulación.

PALABRAS CLAVE: Problema de asignación de turnos de personal (ETP), aprendizaje automático (Machine learning), Q-Learning, forecasting, series de tiempo.

ABSTRACT

This research addresses the problem of assigning personnel shifts at a service station (EDS) called Transpiedecuesta SA. The proposed solution consists of two phases. In the first phase, a machine learning model is used to model fuel demand, thereby forecasting hour-by-hour staffing requirements. The second phase involves implementing a Q-Learning reinforcement learning algorithm in Python, training an agent with a reward function and policies to efficiently assign employees to different shifts over 3 months. The results show an assignment schedule without fixed shifts, with adequate metrics for approximately 16 employees and a total assignment cost lower than the average of the company's actual payroll. In terms of computational cost, 25 minutes and a maximum of 1.5 GB of RAM were used to perform 1200 simulation episodes.

KEY WORDS: Personnel shift assignment problem (ETP), machine learning, Q-Learning, forecasting, time series.

1. INTRODUCCIÓN

A raíz de los retos de deserción laboral, las organizaciones se enfrentan a nuevos desafíos relacionados al talento humano, debido a que las empresas con alta rotación de personal son menos eficientes y eficaces en comparación a aquellas que si mejoran su retención del recurso humano [1].

Actualmente las metodologías que usan las empresas para definir cronogramas de trabajo, asignación de tareas, definir grupos de personal y recursos son muy básicas (y por lo general empíricos), además de no tener un enfoque

[1] Trabajo de grado presentado para optar al título de Ingeniero Industrial en modalidad de artículo científico dirigido por Néstor Raúl Ortiz Pimiento, PhD. en Ingeniería. Universidad Industrial de Santander. Facultad de Ingenierías Fisicomecánicas. Escuela de Estudios Industriales y Empresariales. 2024.

científico, no contemplan realmente las variables que caracterizan el entorno laboral, un entorno siempre cambiante e impredecible que incluye muchas variables para su entendimiento. Las listas de personal, cronogramas y rutas de trabajos que se definan, no sólo deben satisfacer los objetivos empresariales básicos como lo son la minimización de los costos, satisfacción de los clientes y la demanda, sino también deben ajustarse a las políticas, normativas laborales que se estipulan por ley, exigencias de los mismos trabajadores y al comportamiento del mismo entorno.

Todo esto hace que modelar un problema de asignación de tareas a un conjunto de personas en horarios y periodos definidos que cumplan todas las restricciones, sea un proceso complicado e implique metodologías complejas para su solución. Actualmente los avances en ciencias de computación han brindado herramientas como las metodologías de aprendizaje automático o Machine Learning (ML por sus siglas en inglés) para solucionar problemas en ambientes estocásticos como lo es el aprendizaje reforzado. En la literatura hay muy poca evidencia del empleo de estas técnicas en problemas de asignación de turnos de personal o Employee Timetabling Problem (ETP por sus siglas en inglés). Por lo que esta investigación pretende adaptar una metodología novedosa y efectiva para solucionar problemas de asignación que se adecuen a procesos estocásticos como es el caso de la estación de servicio automotriz Transpiedecuesta SA.

2. REVISIÓN DE LITERATURA

Para que la ecuación de búsqueda fuera lo más completa posible se definieron 3 grupos de palabras clave relacionadas con el tema de investigación. En la Tabla 1 se muestran las 3 categorías donde cada una abarca un conjunto de sinónimos.

Tabla 1. Palabras clave de la ecuación de búsqueda

Categoría	Inglés	Español
Acción	Scheduling	Programación
	Assignment	Asignación
	Allocation	Asignación
	Assignation	Asignación
	Programming	Programación
	Programing	Programación
	Planning	Planeación
Objeto asignado	Rostering	Lista
	Timetabling	Horario
	Schedule	Calendario
	List	Lista
	Roster	Lista
	Timetable	Horario
Recurso	Employee	Empleado
	Staff	Personal
	Personnel	Personal
	Workforce	Trabajador
	Operators	Operario

Fuente. Elaboración propia.

El problema de asignación de turnos de personal es una variante específica de problemas más generales como se muestra en el Figura 1.

Figura 1. Variantes del problema de asignación de turnos de persona **Fuente**. Elaboración propia.
Nota. Está figura muestra los problemas generales de los cuales se deriva el problema ETP.

Por eso se definieron 9 ecuaciones de búsqueda en donde se hace un análisis bibliométrico desde el tema macro (o general) hasta el tema específico. Para generar las ecuaciones se hicieron combinaciones entre las tres categorías de palabras (Tabla 1) y luego de traducir las frases resultantes se descartaron las traducciones no válidas. Los resultados arrojados por Scopus se resumen en la Tabla 2 y se analizaron a través de la herramienta Bibliometrix (comando de R).

Tabla 2. Cantidad de documentos encontrados por ecuaciones de búsqueda

Ecuación	Tema de búsqueda	Resultados
1	Problema de asignación	11253
2	Problema de asignación e investigación de operaciones	360
3	Problema de asignación y optimización combinatoria	3890
4	Problema de asignación y Machine Learning	110
5	Optimización combinatoria y Machine Learning	408
6	Problema de asignación, optimización combinatoria y Machine Learning	46
7	Problema de asignación de turnos de personal	1477

Ecuación	Tema de búsqueda	Resultados
8	Problema de asignación de turnos de personal y Machine Learning	47
9	Problema de asignación de turnos de personal y Aprendizaje reforzado	16

Fuente. Elaboración propia.

Los resultados de las 6 primeras ecuaciones muestran que en los últimos años la investigación en problemas de asignación (problema general de asignación) y los relacionados con lo optimización combinatoria ha tomado una gran popularidad entre los investigadores, sin embargo, se evidencia el poco uso de algoritmos de aprendizaje automático en este tipo de problemas. La ecuación de búsqueda 7 muestra el problema general de asignación de turnos de personal (ETP) que se trata en esta investigación. Ya que el número de investigaciones es muy elevado, se aplican los siguientes filtros:

- Año de publicación: 2010 a 2022
- Áreas: Ciencias de la computación, Ciencia de la decisión, Ingeniería, Matemáticas, Administración y Finanzas.

A pesar de que los resultados están filtrados desde el año 2010 se considera importante destacar investigaciones anteriores dado sus grandes aportes. El autor Amnon Meisels, da un punto de partida con sus investigaciones desde 1996 hasta 2003 en donde aborda específicamente el problema ETP y le da un enfoque de red de restricciones (CN). Luego se destacan 2 artículos publicados en el 2004 por A.T. Ernst et al. En el primer documento titulado Staff scheduling and rostering: A review of applications, methods and models realiza una revisión de literatura muy completa y detallada del problema general ETP en donde presenta el proceso de asignación de turnos como una serie de módulos secuenciales, resume las áreas de aplicación del problema y los métodos de solución empleados hasta la época. Sin embargo, la planificación de la fuerza de trabajo y el impacto sociológico y psicológico de diferentes patrones de trabajo fueron 2 áreas relacionadas que no se trataron en la revisión. En su segunda publicación titulada An Annotated Bibliography of Personnel Scheduling and Rostering realiza un compendio de 700 artículos, centrándose en algoritmos para generar listas y horarios de personal, donde cubre áreas relacionadas con la planificación de fuerza laboral y la estimación de requisitos de personal. En esta revisión no incluye documentos relacionados con el proceso de asignación de turnos.

Estos dos autores fueron de gran importancia ya que muchas investigaciones posteriores la referenciarían tal como muestra la red de co-citaciones de la ecuación 7, en la Figura 2.

Posteriormente en el 2013 Jorne Van den Bergh et al. realizan una nueva revisión de literatura sobre el problema ETP titulada Personnel scheduling: A literature review, en este recopila a lo largo del tiempo aportes de diferentes autores desde la formulación de Dantzing y Edie en 1954 hasta la clasificación del problema ETP de Alfares, Ernst y Burke et al. (entre otros) en 2004. Bergh concluye en esta revisión:

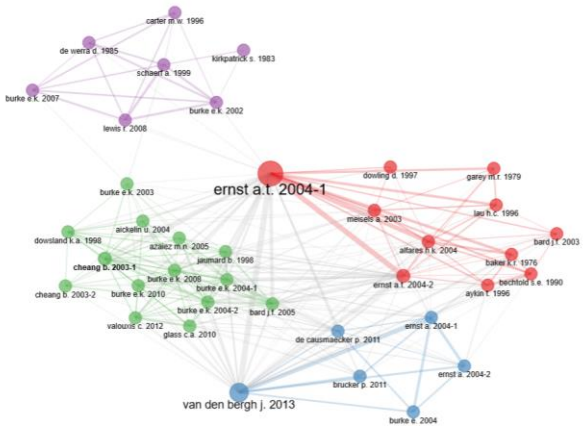

Figura 2. Análisis de cluster. Adaptado de Bibliometrix
Fuente. Elaboración propia.

"… Con respecto al método de solución considerado, observamos que la literatura se inclina en gran medida hacia los enfoques de programación matemática y metaheurísticos. Los algoritmos de descomposición y las técnicas híbridas reciben cada vez más atención por parte de los investigadores, que intentan abordar el problema de la programación de personal con grandes restricciones" (p. 14)

Hasta este punto se puede resumir los métodos de solución y las áreas de aplicación del problema ETP en la Figura 3.

Otro aporte importante de Bergh et al. es que resalta la importancia de la incorporación de la incertidumbre en el planteamiento de problemas ETP. De hecho, clasifica el compendio de artículos estudiados en 2 grandes categorías: determinísticos y estocásticos. Dentro de esta segunda categoría menciona que la incertidumbre se puede dar ya sea en la demanda, en la tasa de llegada o en la capacidad. Van den Bergh señala:

(…) En situaciones en las que la incertidumbre tiene un fuerte efecto en la programación del personal, como la volatilidad de la demanda o los cambios de última hora, podría resultar muy beneficioso incorporar esta

incertidumbre en el proceso de toma de decisiones. (p. 14)

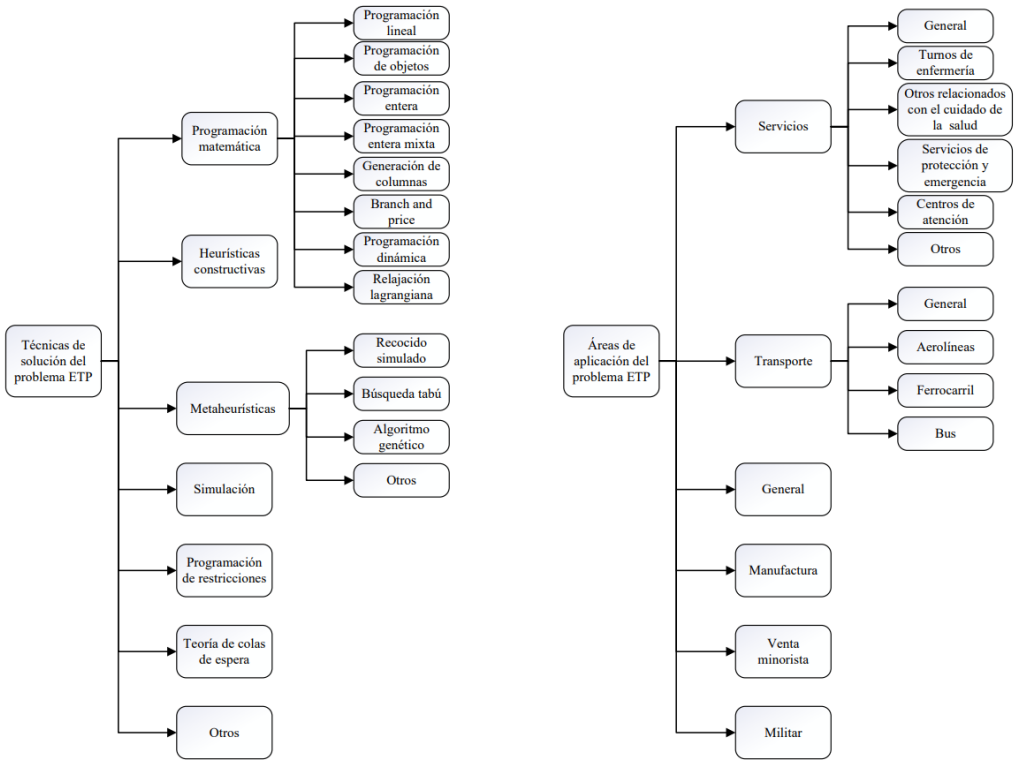

Figura 3. Técnicas de solución y Áreas de aplicación del problema ETP **Fuente**. Adaptado de Personnel scheduling: A literature review (p. 377 y p. 379) por Van den Bergh et al., 2013, European Journal of Operational Research.

Posterior a esto en cuanto a técnicas y algoritmos de aprendizaje automático aplicados a la solución de problemas ETP se encontraron muy pocas investigaciones. Respecto a los resultados de la ecuación de búsqueda 8 sólo fueron útiles 20 publicaciones entre 1998 y 2022, y de la ecuación 9 se seleccionan 15 documentos publicados entre los años 2003 y 2022.

Basados en esto se evidencia que, dentro del Aprendizaje automático, a pesar de la poca investigación en el tema el aprendizaje reforzado ha tomado fuerza para solucionar problemas relacionados con la asignación de turnos de trabajo en los últimos años. Se observa que una gran parte de estas aplicaciones se ha presentado a partir de la última década y se han dado en programación de horarios de exámenes y horarios de cursos (aplicaciones en centros educativos). En otros campos de aplicación, por ejemplo, en el 2020 Matos G.P. et al en su investigación titulada Solving periodic timetabling problems with SAT and machine learning abordan la programación periódica de horarios de transporte público con respecto a los tiempos de viaje utilizando enfoques de problemas de satisfacción booleana (SAT) y de aprendizaje por refuerzo con multiagentes, donde señalan que esta combinación de técnicas nunca se había aplicado en este campo hasta ese entonces.

Por otro lado, una investigación realizada en el 2010 por Guo F. y Song H en Wuhan, China, aborda 2 técnicas de aprendizaje automático como base para el modelo de programación de horarios: aprendizaje reforzado y clasificador Naive Bayes. Esta investigación es lo más cercano que se pudo encontrar al caso que se aborda en esta investigación. En el documento titulado Research and application of data-mining technique in timetable scheduling señalan:

(…) presenta el modelo de programación de horarios basado en el algoritmo de aprendizaje por refuerzo, que puede resolver nuevos problemas de programación de horarios mediante la modificación del vector propio de horarios y del vector de acción de programación de horarios. Al mismo tiempo, se diseña e implementa un sistema de extracción de datos históricos de horarios basado en el algoritmo de clasificación Naive Bayesian. El resultado muestra que la base de conocimientos para la programación de horarios puede construirse rápida y eficazmente mediante el algoritmo de clasificación bayesiano ingenuo [16]

Finalmente se concluye que la mayoría de las aplicaciones de aprendizaje reforzado se ha dado en campos muy específicos de asignación de horarios y

turnos de personal (la gran mayoría horarios de clases), pero no se ha abordado o estudiado de manera general en el problema ETP y más específicamente no se encontraron investigaciones donde empleen el algoritmo Q-Learnig como método de solución para este tipo de problemas. A pesar de que con la ecuación de búsqueda 9 se encontraron aplicaciones de algoritmos de aprendizaje reforzado en problemas de asignación de turnos, estos abordan variantes muy específicas de este problema como por ejemplo construcción de horarios de trenes de diferentes sistemas ferroviarios (Li W., Ni S. 2022), optimización de los horarios del transporte público con respecto a los tiempos de viaje (Matos G.P., Albino L.M., Saldanha R.L., Morgado E.M. 2021) o programación de exámenes a estudiantes en instituciones educativas (Han, Kehan, 2018). Estos problemas, en su mayoría, emplean múltiples agentes de aprendizaje además de que crean algoritmos híbridos combinando con metaheurísticas. Puesto que los espacios de asignación, las acciones, las variables y políticas del entorno de los casos anteriormente mencionados difieren significativamente con los que se definen para el caso de la EDS Transpiedecuesta, nos es posible hacer una adaptación de estas metodologías. Mientras que en un problema de programación de exámenes (problema más común en la revisión) se busca asignar exámenes académicos de forma que no se crucen e interfiera con otros de acuerdo a la disponibilidad de salones y de horarios de docentes y estudiantes, en el problema de asignación de turnos en la EDS se busca asignar un conjunto de operarios a turnos de forma que se cumpla con una normatividad laboral y una demanda de producto en puntos físicos de despacho; los parámetros y variables no son comparables.

3. CONTEXTO DEL PROBLEMA

La estación de servicio Terpel Transpiedecuesta es una estación de servicio automotriz perteneciente a la empresa Transportes Piedecuesta SA ubicada en el kilómetro 2 vía Bogotá, en Piedecuesta Santander. La EDS está distribuida en un terreno de aproximadamente 2900 m2 y cuenta con puntos de distribución de combustible tal y como se muestra en la Figura 4.

Como se observa, la estación de servicio está distribuida físicamente en islas, equipos y mangueras. Cada manguera surte un único tipo de combustible de los tres que se venden (Bio-Diesel o ACPM, gasolina corriente y gasolina extra), a un solo vehículo a la vez y maniobrada por un solo operario (empleado o islero). En cada isla, el proceso se comporta como un sistema de colas con múltiples servidores, en donde los vehículos llegan a la manguera disponible más inmediata, son atendidos por un operario y luego abandonan la estación. La distribución de los puntos de despacho se especifica en la Tabla 3 (por ejemplo, la isla 4 contiene a los equipos 7 y 8; el equipo 7 tiene 2 mangueras que despachan gasolina corriente y el equipo 8 tiene 4 mangueras: 2 que despachan gasolina corriente y 2 gasolina extra).

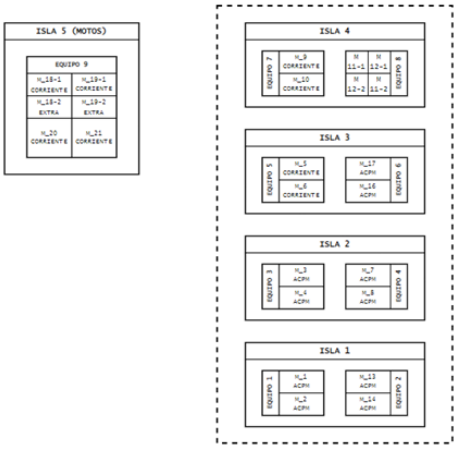

Figura 4. Distribución física de la EDS Transpiedecuesta SA
Fuente. Elaboración propia.

Tabla 3. Relación islas, equipos y tipo de combustible en la EDS

Id_Isla	Id_Equipo	Cantidad de mangueras	Combustible
1	1	2	ACPM
1	2	2	ACPM
2	3	2	ACPM
2	4	2	ACPM
3	5	2	Gasolina corriente
3	6	2	ACPM
4	7	2	Gasolina corriente
4	8	2	Gasolina corriente
4	8	2	Gasolina extra
5	9	4	Gasolina corriente
5	9	2	Gasolina extra

Fuente. Elaboración propia.

En este contexto es de interés definir el cronograma de turnos de trabajo para cada empleado y luego asignarlos en los distintos puntos de despacho de combustible de manera que se minimicen los costos asociados a dicha asignación, se atienda la demanda de combustible de los vehículos y se cumpla con la normatividad laboral vigente.

La mínima unidad de tiempo que se empleará será la hora y a nivel de puntos de despacho se hará por equipos, lo que quiere decir que el problema consistirá en asignar operarios a equipos en cada hora del día.

4. PROBLEMA DE INVESTIGACIÓN

En la mayoría de los problemas ETP convencionales que se resuelven con programación lineal entera, la variable de modelamiento sólo indica cuántos empleados deben ser asignados a cada turno en un periodo de planeación definido. En estos problemas se asume que se conoce la duración de los turnos y el requerimiento de personal para cada unidad de tiempo, los cuales no dependen de una demanda externa. Adicional a esto, es difícil hacer un seguimiento riguroso a cada empleado por individual, en el sentido de que la rotación que se debe hacer al personal de forma que cumplan con ciertas horas laboradas y de descanso a la semana o al mes es una tarea adicional que debe solucionar la empresa o el investigador; el algoritmo dice cuántos, pero no quiénes ni de qué manera.

Las investigaciones y publicaciones consultadas en el análisis de literatura se basan en las características y supuestos anteriormente mencionados. En las investigaciones tituladas Employees Timetabling Simulation using Integer Linear Programming Technique (Sudradjat Supian, R. Sudrajat, Eman Lesmana y Amalia Farida, 2017) y A Polyhedral Approach for the Staff Rostering Problem (Giovanni Felici, Claudio Gentile, 2004), abordan un problema de asignación de turnos de personal teniendo en cuenta políticas corporativas y necesidades de personal en un problema de optimización lineal entera binaria. Una particularidad en la metodología de estas dos investigaciones es que se asigna una lista de empleados inicial a través de una matriz binaria para cada día y turno, por lo que es posible hacer seguimiento de la jornada laboral a cada empleado; la metodología abordada en dichas publicaciones es la que presenta mayor adaptación con lo que se pretende lograr en la EDS Transpiedecuesta, sin embargo, al igual que el resto de investigaciones, se parte del hecho de que se conocen de antemano los turnos y necesidades de personal, además de que no se relacionan dichas cantidades con una demanda de producto, un puesto físico de trabajo rotativo (a excepción de las investigaciones que integran el problema ETP con sistemas de producción Job-shop),una distribución de carga laboral homogénea y una retribución equitativa (comisiones por venta).

Al adaptar el problema a la actual situación de la EDS hay que hacer una doble asignación de los empleados, por lo que es muy complejo abordar el problema con un modelo de programación lineal entera, ya que al asignar el personal a turnos se busca minimizar los costos (minimizar la cantidad de personal empleado para atender la demanda) y al asignarlos a los punto físicos de despacho se busca minimizar la variabilidad en la carga laboral que se distribuye entre ellos (los costos de asignación depende del turno, no del lugar); lo anterior implicaría plantear un modelo multi-objetivo con un gran número de parámetros, variables y restricciones complejas. En la metodología propuesta en esta investigación se divide el problema en tres problemas de menor complejidad, en donde no se tienen definidos turnos fijos, sino que un agente de aprendizaje va asignando empleados hora a hora de una lista de disponibles en un entorno dinámico y de acuerdo con políticas y restricciones del entorno. Es así que la gran diferencia con los métodos convencionales es que, bajo este enfoque, a medida que se va asignando empleados, el agente evalúa en cada momento si es necesario o no que alguien "nuevo" sea contratado para atender la demanda, de esta manera el algoritmo empleará la menor cantidad posible de personal para hacer cumplir los requerimientos. Otra gran ventaja es que el algoritmo hace la rotación de cada empleado dentro del periodo de planeación de forma que cumpla con la normatividad laboral vigente y con las políticas laborales de la empresa; de esta manera es posible hacer un seguimiento a cada empleado en el sentido de que es posible visualizar sus cronogramas de trabajo y descanso, ya que el algoritmo reconoce a cada empleado como único y diferente dentro del grupo.

5. METODOLOGÍA DE SOLUCIÓN

Dado que es un problema de doble asignación (empleados a turnos y a zonas de despacho) y que el requerimiento de personal para cada hora y equipo depende de la demanda de combustible (asumiendo una relación proporcional), se decide segmentar la solución del problema en dos fases para la construcción de los calendarios de turnos de personal

- Fase 1: Modelamiento de la demanda (forecasting) y cálculo de requerimiento de personal
- Fase 2: Asignación de empleados a turnos a través de Q-Learning

En la primera fase se modela y predice la demanda de combustible para cada punto de despacho hora a hora a través de series de tiempo y algoritmos de Machine Learning, de esta manera se obtiene un estimado del tiempo de ocupación y el requerimiento de personal para una ventana de tiempo de tres meses (tercer trimestre de 2023). Posteriormente en la fase 2, el agente de aprendizaje asignará de manera secuencial cada empleado a cada hora del día de forma que se cumpla con el requerimiento de personal y las políticas laborales de la empresa. Para dicha asignación el agente iterará en cada hora, de acuerdo con una función de recompensa y una matriz de estados; al incumplir determinadas políticas o restricciones del entorno, se le "penalizará" con valores moderadas o graves de acuerdo con el grado de incumplimiento. Por último, se aclara que, puesto que cada tipo de asignación busca objetivos distintos, las 2 fases propuestas sólo abordan la asignación de los

operarios a turnos de trabajo, la segunda asignación a puntos físicos de despacho se propone como una continuación a esta investigación.

5.1. Fase 1. Modelamiento de la demanda y cálculo de requerimiento de personal.

Cada vez que una manguera surte algún tipo de combustible a un vehículo, el sistema almacena datos en las siguientes variables:

- $Tipo\ Transaccion$: indica el método de pago del cliente (Credito local, Credito satelite, Efectivo, Prepago local, Privada local)
- $Producto$: Tipo de combustible despachado (A.C.P.M., CORRIENTE, EXTRA)
- $Consecutivo$: Número único e irrepetible que identifica cada transacción (venta)
- $Fecha$: Momento (Año, mes, día, hora, minuto y segundo) en el que se realiza el despacho de combustible
- $Volumen$: Volumen (en Galones) de combustible despachado.
- $Dinero$: Valor del combustible despachado (COP).
- $Posicion$: Código numérico que identifica la manguera que despachó el combustible.
- PPU: Precio unitario del combustible despachado (COP/Gal)
- $Placa$: Placa del vehículo al cual se le despachó el combustible
- $Nombre_cuenta$: Indica la razón social para clientes registrado en la estación.

Para el desarrollo del problema se cuenta con una base de datos con 1.868.889 registros de las anteriores 10 variables, tomados desde el 3 de febrero de 2020 hasta el 30 de junio de 2023. Dicha base de datos es suministrada por la administración de la EDS, con permiso de uso para fines netamente académicos.

5.1.1. Construcción del dataframe.

Teniendo en cuenta que al año 2020 fue atípico dada la pandemia por el COVID-19 se hace un análisis preliminar de las ventas de combustible. En la Figura 5 se evidencia que el comportamiento de las ventas para el 2020 fueron más bajas comparadas con los años posteriores, además que el mes de abril hace que el comportamiento de la serie difiera significativamente de las demás. Por todo esto, se deciden excluir los datos correspondientes al año 2020 ya que se consideran sesgados por un evento externo no controlable.

Inicialmente solo se consideran relevantes las variables $Producto, Fecha, Volumen, Posicion$ y PPU, siendo la variable $Volumen$ la variable principal de análisis (variable dependiente). Como interesa que la mínima unidad de tiempo sea la hora y además que los despachos de combustible se analicen por equipo, se crea una nueva variable denominada $Equipo$ la cual es un código numérico que identifica el equipo al cual pertenece la

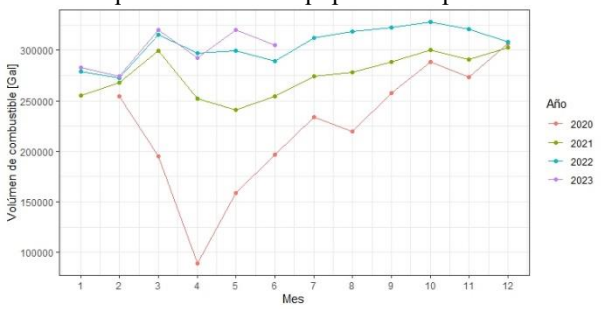

Figura 5. Comportamiento de la demanda de combustible en los últimos 3 años. **Fuente**. Elaboración propia.

manguera (dicha variable reemplaza la variable $Posicion$), se crean nuevas variables (a partir de la variable $Fecha$) que corresponden a distintas agrupaciones de tiempo y se agrupan los datos, obteniendo finalmente un dataframe de 177.029 observaciones con 16 variables (Tabla 4).

Tabla 4. Variables que conforman el dataframe.

Id	Variable	Definición	Tipo de dato
1	Volumen	Volúmen de combustible despachado (Gal)	Numeric
2	Tanqueos	Cantidad de tanqueos realizados (vehículos atendidos)	Integer
3	Producto	Tipo de combustible despachado	Character
4	Precio Galon	Precio unitario del combustible despachado (COP/Gal)	Numeric
5	Equipo	Código numérico que identifica al equipo del que despachó el combustible	Integer
6	Isla	Isla en la que se despchó el combustible	Integer
7	Fecha4	Momento (Año, mes, día y hora) en el que se realiza el despacho de combustible	Date
8	Fecha2	Momento (Año, mes y día) en el que se realiza el despacho de combustible	Date
9	Hora	Hora del día en la que se realizó el despacho	Integer

Id	Variable	Definición	Tipo de dato
10	Dia Semana	Día de la semana en el que se realizó el despacho	Integer
11	Dia Mes	Día del mes en el que se realizó el despacho	Integer
12	Dia Anio	Día del año en el que se realizó el despacho	Integer
13	Semana Anio	Semana del año en el que se realizó el despacho	Integer
14	Mes	Mes en el que se realizó el despacho	Integer
15	Anio	Año en el que se realizó el despacho	Integer
16	Festivo	Indicador binario si el día del despacho es feriado	Integer

Fuente. Elaboración propia.

Luego de la selección de variables, agrupación de datos y definición de nuevas variables, la dimensionalidad de los datos se redujo en un 85%. Con este último dataframe se estudiará el comportamiento de las variables $Volumen$ y $Tanqueos$ (las principales variables de interés) en función de las 14 variables restantes.

5.1.2. Análisis exploratorio de datos (EDA)

En un primer análisis, la Figura 6 muestra que la distribución de la cantidad de combustible despachado (en Galones/hora*equipo) tiene un gran sesgo positivo o hacia la derecha, lo que deriva una gran cantidad de registros atípicos (por arriba). El promedio de despachos de combustible está en 49,479 Gal/h*equipo en donde el 80% de ellos no supera los 72,2 Gal/h*equipo, sin embargo, se registran casos en donde un equipo llega a despachar más de 800 Gal/h. La razón por la que se presentan tantos datos atípicos se justifica más adelante cuando se realice el análisis por islas y por equipos.

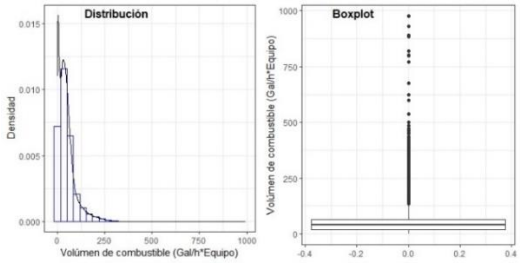

Figura 6. Distribución de la demanda de combustible (Galones/hora*equipo) **Fuente**. Elaboración propia.

En cuanto al arribo de vehículos por hora a cada uno de los equipos, la Figura 7 se evidencia también un sesgo positivo o hacia la derecha en la distribución de la variable, lo que implica una gran cantidad de valores atípicos (muy altos). En estadísticas descriptivas el valor mínimo es de 1, que indica que en cualquier hora arriba por lo menos un vehículo a la estación; en el 50% de los casos se realizan como máximo 5 tanqueos, y en el 80% de los casos 12 tanqueos como máximo. Se presentan casos donde un equipo realiza más de 80 despachos de combustible en una hora.

Al analizar las dos variables principales en cada una de las islas, se observa que las islas 1 y 2 son las que mayor volumen de combustible despachan por hora y por equipo, y las islas 4 y 5 las de menor volumen de despacho. Pero este comportamiento es contrario al observar la frecuencia con la que llegan los vehículos a las islas: las islas 1 y 2 son las que tienen menor cantidad de despachos por hora y por equipo, y las islas 4 y 5 presentan la mayor cantidad (Figura 8). Lo anterior da un indicio de que efectivamente existen diferencias en cuanto a volumen y cantidad de despachos entre islas. Por último, se agrega que los valores atípicos en las dos variables se presentan en todas las islas.

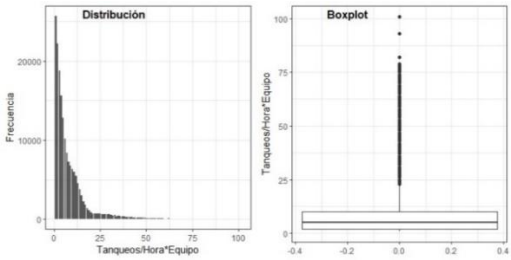

Figura 7. Distribución de la tasa de llegada de vehículos por hora. **Fuente**. Elaboración propia.

La causa por la que entre islas se presentan diferencias en cuanto a volumen de combustible y cantidad de despacho se hace más evidente cuando analizamos estas dos variables por equipos. En la Figura 9 se observa que el equipo 2 tiene mayores volúmenes de combustible despachado por hora, pero es en este dónde se presenta las menores cantidades de despacho (tanqueos); caso contrario ocurre con el equipo 9, en donde se presentan los menores volúmenes de combustible despachado por hora, pero presenta una frecuencia de despachos mucho más alta que la de los demás equipos.

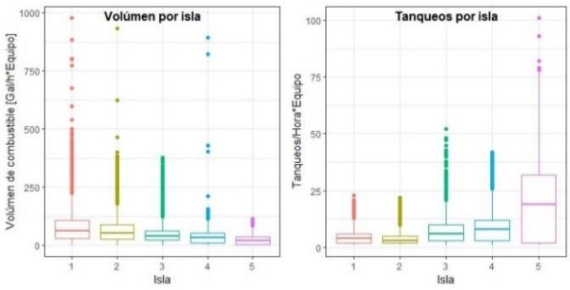

Figura 8. Distribución de las dos variables principales por isla. **Fuente**. Elaboración propia.

Las figuras 8, 9 y10 muestran que en los equipos que se despachan ACPM, tienen mayor demanda de volumen de combustible, pero menor frecuencia de despacho, en tanto que los equipos que despachan gasolina (corriente o extra) presentan menor volumen de combustible despachado, pero mayor frecuencia de cantidad de tanqueos. Estos resultados concuerdan con la situación real de la estación de servicio, pues los vehículos que demandan ACPM son vehículos medianos y pesados, vehículos robustos con una gran capacidad de tanqueo como lo son camiones, buses, tractomulas, etc.

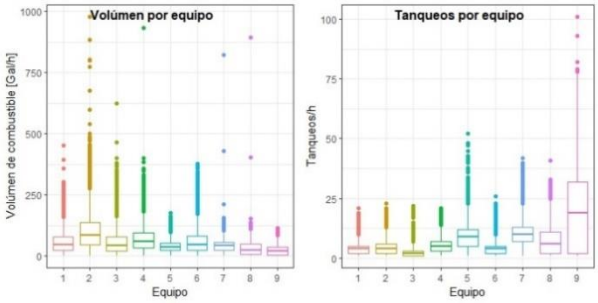

Figura 9. Distribución de las dos variables principales por equipo. **Fuente**. Elaboración propia.

Este tipo de vehículo tardan mayor tiempo en ser atendidos lo que hace que la frecuencia de llegada por hora disminuya. Por otro lado, están los vehículos que demandan gasolina corriente o extra, estos son de tipo liviano y motocicletas, vehículos que llegan con mucha más frecuencia a la estación, pero dada su capacidad de almacenamiento de combustible tardan menos en ser atendidos. Ejemplo de esto es la isla 9 con su único equipo (equipo 9), al ser una isla exclusivamente para motocicletas, estas llegan en mayor cantidad y tienen un tiempo de atención muy reducido.

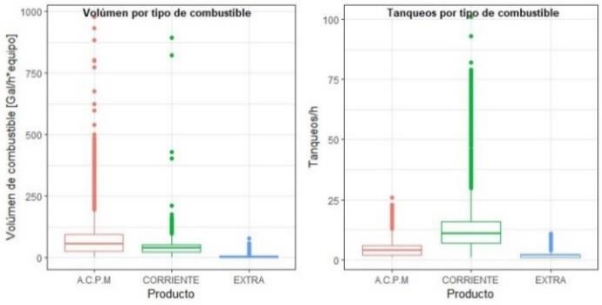

Figura 10. Distribución de las dos variables principales por tipo de combustible. **Fuente**. Elaboración propia.

5.1.3. Definición y modelamiento de series de tiempo

Para la construcción de un horario de turnos de personal en donde la demanda del producto determine el requerimiento de personal hora a hora, es necesario primero modelar el comportamiento de la demanda de volumen de combustible y la cantidad de vehículos a atender en cada uno de los equipos. Recordemos que una serie de tiempo se descompone básicamente en 4 componentes: tendencia, ciclo, estacionalidad y ruido; para controlar dichos componentes en la demanda de combustible y la cantidad de despachos en la EDS, se analiza el comportamiento de estas dos variables agrupándolas por distintas unidades de tiempo, tal y como se detalla en la Tabla 4: hora del día (1-24), día de la semana (1-7), día del mes (1-31), día del año (1-365), semana del año (1-53), mes del año (1-12), año (2021-2023) y además una variable binaria que indica si el día es feriado o no. Justamente estas agrupaciones dan lugar a variables exógenas que posteriormente intentarán justificar las variaciones en los modelos.

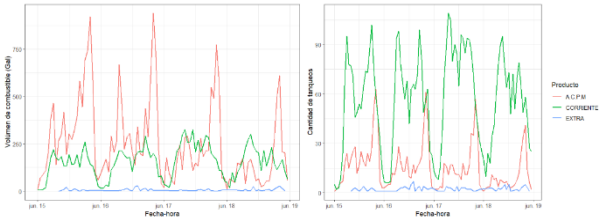

Figura 11. Comportamiento de la demanda por hora y por tipo de combustible en una muestra de 4 días consecutivos. **Fuente**. Elaboración propia.

Un ejemplo de lo anterior se muestra en la Figura 11, en donde la demanda de combustible y la frecuencia de despacho se agrupan por hora en una muestra de 4 días consecutivos. Se observa una estacionalidad que se repite a diario especialmente en el ACPM y gasolina corriente: valores de demanda y arribo de vehículos mínimos al finalizar el día (12 de la madrugada) y valores máximos alrededor de 2 horas específicas que son las 6 de la mañana y las 6 de la tarde.

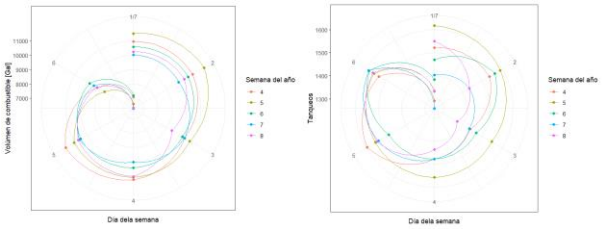

Figura 12. Comportamiento de la demanda por día de la semana en una muestra de 5 semanas consecutivas. **Fuente**. Elaboración propia.

La Figura 12 también muestra que ambas variables presentan una estacionalidad semanal, ya que los valores crecen considerablemente para el día el lunes, se mantienen durante toda la semana y disminuyen los fines de semana. Finalmente, la Figura 13 muestra que el comportamiento de la demanda mensual de combustible durante el primer cuatrimestre es muy idéntico entre los

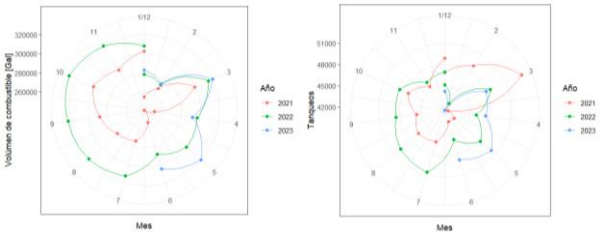

Figura 13. Comportamiento de la demanda por mes desde al año 2021 hasta el año 2023. **Fuente**. Elaboración propia.

tres años, además que se evidencia una tendencia de crecimiento sostenida desde el año 2021 hasta el año 2023.

Todo lo anterior se resume en una descomposición aditiva de la serie de tiempo como se muestra en la Figura 14. Se evidencia que tanto la demanda de combustible como la frecuencia de llegada de los vehículos presentan estacionalidad múltiple y tendencia secular. En estos casos los métodos convencionales de modelamiento de series de tiempo no son muy adecuados, por lo que es necesario recurrir a métodos un poco más complejo para poder abordar el problema.

Puesto que a nivel de puntos de despacho se aborda el problema por equipos, es necesario ahora determinar si hay diferencias entre ellos en cuanto a las dos variables estudiadas. La Figura 9 mostró indicios de que efectivamente existen diferencias, y dado que la cantidad de datos por equipo es muy elevada, cualquier prueba de comparación múltiple encontrará diferencias significativas muy fácilmente con un valor-p muy bajo. Dado que los equipos están agrupados por islas, se procede a comparar los equipos en cada una de ellas, para que, en el caso de no haber diferencias entre los dos equipos, reducir la complejidad del problema asignando empleados por islas y no por equipos (en los que casos que corresponda).

Para realizar la comparación entre dos equipos de una misma isla se observará la distribución muestral a través de una técnica iterativa basada en remuestreo Bootstrap, en donde para cada equipo se seleccionarán de manera aleatoria k muestras de datos tamaño n sin reemplazamiento y se compararán las distribuciones muestrales de la media. Como ejemplo se presenta el caso de la isla 1, en donde la distribución del volumen de combustible despachado por hora para cada equipo se compara en la Figura 15.

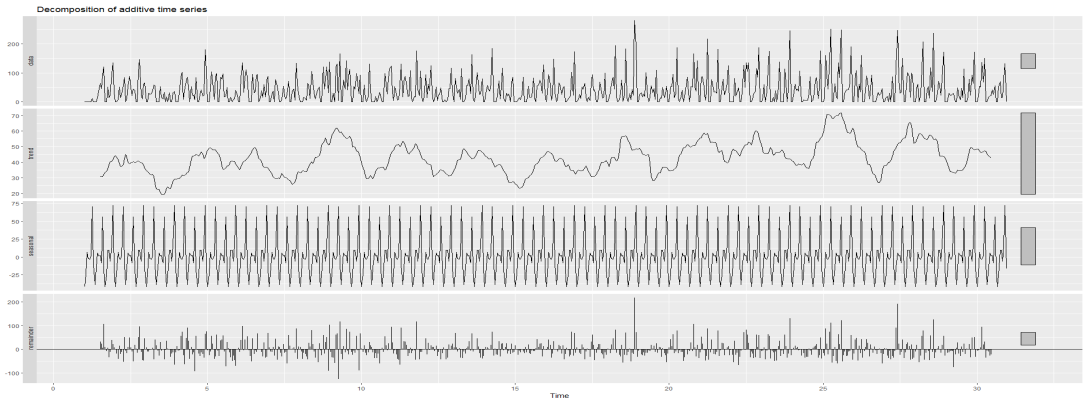

Figura 14. Descomposición aditiva de la serie de tiempo agrupada por hora en una muestra de 30 días. **Fuente**. Elaboración propia.

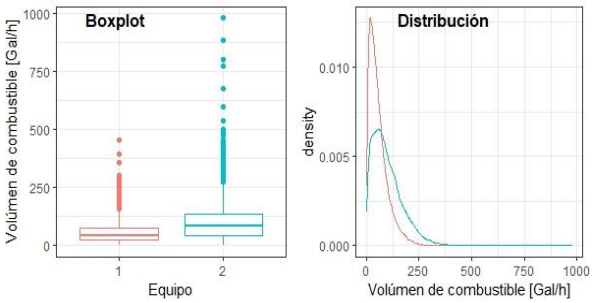

Figura 15. Comparación de las dos variables estudiadas en los dos equipos de la isla 1. **Fuente**. Elaboración propia.

El boxplot de la Figura 15 indica que, a pesar de que los equipos 1 y 2 despachan el mismo tipo de combustible (ACPM) parece que el equipo 2 tiene una mayor demanda. Ambas distribuciones se encuentran sesgadas hacia la derecha y no tienen una distribución normal. Se procede a seleccionar aleatoriamente k = 9999 muestras de cada equipo, todas las muestras de tamaño n = 100 y sin reemplazamiento, y se grafican las distribuciones muestrales.

La Figura 16 muestra una clara diferencia en las dos distribuciones. Mientras que el 95% de las medias muestrales se encuentran entre 48,36 y 66,03 Gal/h para el equipo 1, esta misma proporción se encuentra entre

85,54 y 113,61 Gal/h para el equipo 2, por lo que se concluye que para la isla 1, la demanda del volumen de combustible es distinta entre sus dos equipos. Este mismo análisis se realizó para las 3 islas restantes (para la isla 5 no porque esta solo cuenta con un equipo) obteniendo los mismos resultados (los dos equipos difieren en cuanto volumen de combustible despachado) por lo que se modelarán series de tiempo distintas e independientes para cada equipo.

Para el caso se definen 18 series de tiempo, puesto que son 9 equipos y en cada uno se modelan dos variables distintas (volumen de combustible demandado en galones, y cantidad de vehículos a atender).

Medida	Equipo 1	Equipo 2
Mínimo	42,45	74,68
Q1	53,65	94,31
Q2	56,65	99,08
Media	56,78	99,25
Q3	59,76	103,99
Max	75,49	132,85
P 2,5	48,36	85,54
P 97,5	66,03	113,61

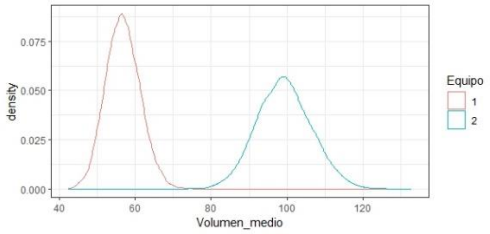

Figura 16. Comparación de las distribuciones muestrales de medias para los dos equipos. **Fuente**. Elaboración propia.

Las técnicas para modelar dichas series de tiempo se dividen en dos grandes categorías: métodos de series de tiempo y algoritmos supervisados de aprendizaje automático. De la primera categoría se proponen 9 tipos de modelos dentro de los cuales se encuentran el método Naive estacional, ARIMA con variables exógenas, STL + ETS(A, N,N), Regresión armónica dinámica con múltiples períodos estacionales y modelos TBATS (modelo de espacio de estado de suavizado exponencial con transformación de Box-Cox, errores ARMA, componentes de tendencia y estacionales); los dos últimos métodos implica el suavizado de componentes a través de series de Fourier. En la segunda categoría se proponen 4 tipos de algoritmos supervisados cada uno probado con dos librerías distintas, para un total de 8 modelos distintos. Dentro de esta segunda categoría se encuentran modelos de árboles de regresión, bosques aleatorios, redes neuronales artificiales feed-forward multicapa, con una sola capa oculta, entradas retrasadas y variables exógenas; para el respectivo entrenamiento se definieron como variables exógenas las propias componentes de tiempo definidas al inicio de la sección 3.4.

Para validar cada uno de los modelos se empleó el método de validación simple, en donde el dataset definido en la Tabla 4 se divide en dos partes, entrenamiento y validación; el primero consta de todos los datos comprendidos entre el 01/01/2021 hasta el 31/03/2023 y el segundo abarca el segundo trimestre de 2023. Con el primer dataset se ajustan y se calculan los mejores parámetros para el modelo y con el segundo se compara la capacidad predictiva del modelo; justamente con los datos de validación se calculan tres métricas de error (Error Medio EM, Error Cuadrático Medio RMSE y Error Medio Absoluto MAE) y la distribución del error de estimación. Para cada una de las 18 series de tiempo se escoge 1 de los 17 modelos ajustados, de acuerdo al que muestre las mejores métricas.

En la Figura 17 se muestra el comportamiento del RMSE para cada grupo de técnicas en cada variable. Se evidencia que los algoritmos basados en Machine Learning son más precisos que los basados en series de tiempo, pues en general presentan menores métricas de error (el comportamiento es similar con el EM y MAE).

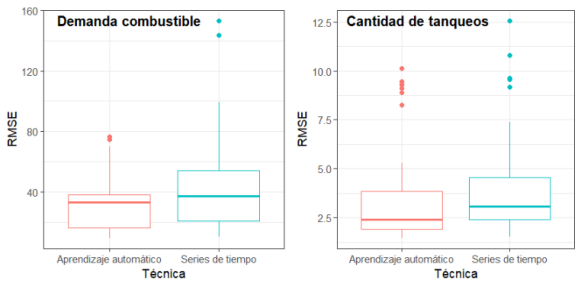

Figura 17. Valores de RMSE por técnicas de modelamiento en las dos variables principales. **Fuente**. Elaboración propia.

En contraposición, la Figura 18 evidencia que a pesar de que la segunda categoría es más precisa en cuanto al pronóstico, implica un coste computacional más elevado. Se aclara que el coste computacional en este estudio abarca desde la búsqueda de los mejores parámetros (entrenamiento) hasta la optimización de hiperparámetros (en el caso de los modelos de Machine Laearning).

A diferencia de las técnicas de la primera categoría, en donde se tuvo que validar cada uno de los 9 métodos en cada una de las 18 series de tiempo, para los algoritmos de aprendizaje automático es posible agregar a cada modelo la variable $Equipo$, con lo que solo basta un solo modelo para modelar la demanda de todos los 9 equipos en simultáneo, lo que aporta una gran ventaja en reducción de la complejidad y el coste computacional (se

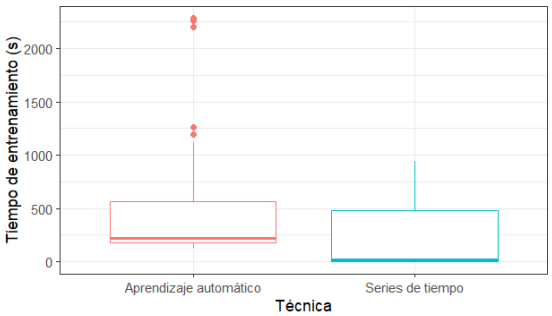

Figura 18. Coste computacional (tiempo) de ajuste del modelo según técnica de modelamiento **Fuente**. Elaboración propia.

comprobó que adicionando la variable $Equipo$ en cada uno de los 8 tipos de modelos, se obtienen métricas de rendimiento similares e incluso mejores que entrenar los 8 modelos en cada una de las 18 series de tiempo de forma independiente).

Del proceso de entrenamiento, validación y selección de modelos, se concluye que, de todos los 17 métodos de modelación propuestos, los bosques aleatorios arrojan las mejores métricas de validación. Los resultados con este algoritmo mostraron también que sólo 4 variables son realmente influyentes tanto en la demanda combustible como en la cantidad de despachos: hora del día, día de la semana, festivo y Equipo.

En conclusión, se emplearán sólo dos modelos de bosques aleatorios (uno para cada variable principal) para pronosticar en todos los equipos y todas las horas. La Figura 19 se muestra como ejemplo para el ajuste de la serie de tiempo de demanda de combustible para el equipo 2 en una muestra de 100 días consecutivos.

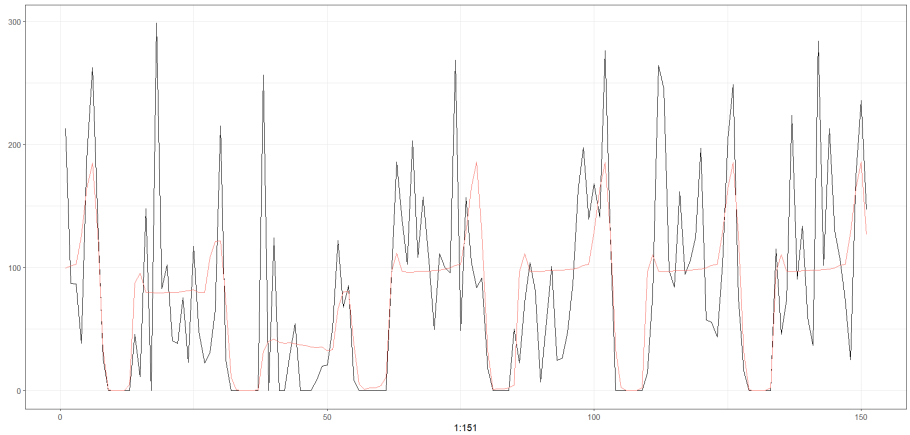

Figura 19. Datos reales (línea negra) y ajustados (línea roja) para la demanda de combustible del equipo 2 en una muestra de 100 días. **Fuente**. Elaboración propia.

5.1.4. Cálculo de requerimiento de personal

Teniendo los pronósticos realizados tanto de la demanda de combustible (Gal) como de la cantidad de despachos (tasa de llegada de vehículos) hora a hora en cada uno de los 9 equipos, se procede a calcular para cada uno de ellos el tiempo de ocupación en cada una de las 2208 horas del horizonte de planeación (92 días x 24 horas/día). El tiempo de ocupación corresponde al tiempo necesario (estimado) que requerirá un operador (Islero) en una hora y equipo específicos para atender la demanda de combustible. Dicho tiempo abarca desde la llegada hasta la salida del vehículo del sistema y se define a través de la Ecuación 1

$$T_{ocp\,ij} = T_{min\,ij} + T_{atn\,ij} \;\; \forall i = 12, \dots, 9; j = 1,2, \dots, 2208 \;\; (1)$$

El tiempo mínimo de despacho ($T_{min\,ij}$) para el equipo i en la hora j es el tiempo que en teoría tardará una manguera en surtir el combustible al vehículo y el tiempo de atención adicional ($T_{atn\,ij}$) es el tiempo adicional que empleará el vehículo en el equipo i en la hora j para ser atendido (incluye tiempos por alistamiento, pago del tanqueo con cierto método de pago, atenciones adicionales, etc). Dado que el tiempo de atención es una variable aleatoria de la cual no se tiene certeza de su distribución y para ello implicaría un estudio de tiempos en cada hora del día lo que introduciría más complejidad al problema, para simplificación se asume una distribución de acuerdo con el criterio de Laplace* y a la experiencia de los despachadores.

El tiempo mínimo de despacho se calcula a través de la Ecuación 2

$$T_{\min ij} = \frac{v_{ij}}{k_i} \;\; \forall i = 12, \dots, 9; j = 1,2, \dots, 2208 \;\; (2)$$

* Criterio de Laplace. El criterio de Laplace parte del hecho de que no se conocen las probabilidades de ocurrencia de cada uno de los estados de la naturaleza, propone que las probabilidades sean las mismas para cada estado.

Donde v_{ij} es al valor pronosticado de la demanda de combustible (Gal) para el equipo i en la hora j y k_i es la capacidad nominal de despacho de una manguera del equipo i (Gal/min).

El tiempo de atención adicional se calcula a través de la Ecuación 3

$$T_{atn\,ij} = c_{ij} * u_{ij} \;\; \forall i = 12, \dots, 9; j = 1,2, \dots, 2208 \;\; (3)$$

Donde c_{ij} es el valor pronosticado de la cantidad de vehículos que llegarán al equipo i en la hora j (redondeado al entero más cercano) y u_{ij} corresponde a un tiempo de atención medio (minutos/vehículo) dado por una variable aleatoria continua con distribución de probabilidad uniforme $u_{ij} \sim U(a_i, b_i)$ donde los parámetros se definen con base a experiencia de los operadores. Los operadores manifiestan que dicho tiempo adicional de atención depende de muchas otras variables como la capacidad del tanque del vehículo, el método de pago, si adquiere o no otros productos de la estación como aditivos y/o lubricantes, etc. Para simplificar definen que para aquellos equipos que surten ACPM dicho tiempo se estima entre 2 y 4 minutos por vehículo, y para los demás está entre los 1 y 3 minutos. Por último, se calcula el tiempo total de ocupación $T_{ocp_tot\,j}$ para cada hora j del día con la Ecuación 4

$$T_{ocp_tot\,j} = \sum_{i=1}^{9} T_{ocp\,ij} \;\; \forall j = 1,2, \dots, 2208 \;\; (4)$$

Para traducir este tiempo estimado de ocupación total por hora en requerimiento de personal se asume una relación lineal y directamente proporcional entre estas dos variables, estableciendo 2 puntos teóricos con los que se puede formar una función lineal. Sea $P_{\min j}$ la cantidad mínima de empleados que se requieren en la hora j para atender la demanda, y además una función que depende de $T_{ocp_tot\,j}$ a través de la Ecuación 5

$$P_{\min j} = \beta_0 + \beta_1 * T_{ocp_tot\,j} \;\; \forall j = 1,2, \dots, 2208 \;\; (5)$$

y que a esta recta pertenecen todos los pares ordenados (t, p). Se asume que, si en un momento dado no llegará ningún vehículo a la EDS, no habrá demanda de combustible y por ende no será necesario asignar un operario, por lo que tenemos el primer par ordenado que pertenece a la recta $P_1 = (0; 0)$ y por ende el valor del primer parámetro $\beta_0 = 0$. Por otro lado, las políticas de la empresa establecen que en cualquier hora j cada equipo tiene que tener asignado sólo un operario (lo que implica que un operario puede estar asignado a varios equipos en una hora dada) por lo que el número máximo de empleados operando en una hora dada debe ser máximo de 9; observando los datos calculados de la variable $T_{ocp_tot\,j}$ se tiene un valor máximo de 431,14 minutos por lo que podemos establecer un segundo punto que conforma esta recta $P_2 = (440; 9)$ (se ajusta los 431,14 a 440) y así podemos calcular el segundo parámetro β_1 de la siguiente manera

$$\beta_1 = \frac{\Delta p}{\Delta t} = \frac{p_2 - p_1}{t_2 - t_1} = \frac{9-0}{440-0} = 0{,}02045 \;\; (6)$$

Como se requieren valores enteros, la expresión se redondea al entero inmediatamente mayor y el modelo final se muestra en la Ecuación 7

$$P_{\min j} = \lceil 0{,}02045 * T_{ocp_tot\,j} \rceil \;\; \forall j = 1,2, \dots, 2208 \;\; (7)$$

Así pues, se finaliza esta primera fase en donde a partir del modelamiento de la demanda, se calcula el requerimiento de personal para cada hora del horizonte de planeación. Estos datos serán una de las entradas principales para programar las restricciones en la siguiente fase. En la Figura 20 se muestra el comportamiento de requerimiento de personal para una muestra de 1 semana. Dicho comportamiento tiene coherencia con los gráficos de demanda estudiados en secciones anteriores. Se observa también que el valor mínimo de $P_{\min j}$ es de 1 el cual se da en horas muy cercanas a la media noche, lo que implica que en estos casos ese solo operario será asignado a todos los 9 equipos dada la baja demanda y además se da cumplimiento a una de las políticas que indica que en ningún momento la EDS se debe quedar sin operario asignados; por otro lado, los picos de requerimiento ocurren alrededor de las 6am y las 6pm. Se fija para cada hora el requerimiento de personal máximo en 9 operarios recordando la política de la empresa.

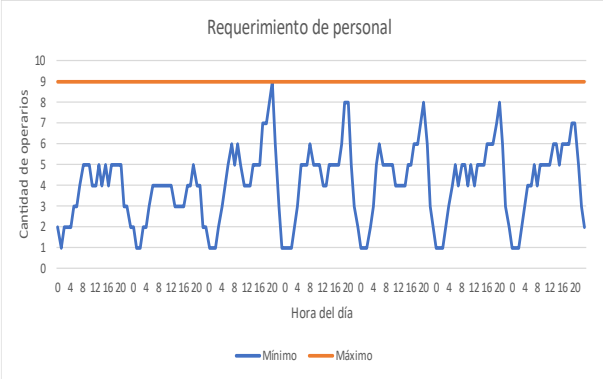

Figura 20. Comportamiento del requerimiento de personal por hora para una muestra de 7 días. **Fuente**. Elaboración propia.

5.2. Fase 2. Asignación de empleados a turnos a través de Q-Learning.

El Q-Learning es uno de los algoritmos más utilizados para solucionar problemas de aprendizaje por refuerzo a

partir de la recompensa retardada. Está basado en Procesos de Decisión de Markov (MDP: Markov Decision Process) y fue introducido por Watkins en 1989 y por Watkins & Dayan en 1992 [2]. "Equivale a un método incremental para la programación dinámica que impone demandas computacionales limitadas" [3]

Para un MDP discreto (X, A, P, r), donde X es el espacio de estados (finito), A es el espacio de acción (finito), P representa las probabilidades de transición y r representa la función de recompensa (FS Melo, 2001), se define el algoritmo Q-Learning como un algoritmo de control de TD fuera de política (off-policy TD control) dado por la Ecuación 7 [4].

$$Q_{t+1}(x_t, a_t) = Q_t(x_t, a_t) + \alpha_t(x_t, a_t)[r_t + \gamma \max_{b \in A} Q_t(x_{t+1}, b) - Q_t(x_t, a_t)] \quad (7)$$

[5]

Donde $Q_t(x_t, a_t)$ denota el valor actual, $\alpha_t(x_t, a_t)$ la tasa de aprendizaje que regula la velocidad en la que se aprende, r_t la recompensa, γ la tasa de descuento, la cual tendrá en cuenta la recompensa a corto o largo plazo, y $\max_{b \in A} Q_t(x_{t+1}, b)$ el valor óptimo esperado [2]

"El algoritmo Q-learning encuentra una política óptima en el sentido de que maximiza el valor esperado de la recompensa total mejorando sucesivamente sus evaluaciones de la calidad de acciones particulares en estados particulares" [3] empezando desde el actual [5]

La construcción del algoritmo consta básicamente de 7 pasos o etapas que se detallan a continuación

5.2.1. Definición del entorno

El entorno de ejecución corresponde a todas aquellos elementos y restricciones que el agente de aprendizaje debe tener en cuenta para realizar sus acciones cuando se encuentre en cada uno de los estados definidos. Dichos elementos corresponden variables, parámetros, hiperparámetros, listas y funciones secundarias las cuales se detallan en la Tabla 5.

Tabla 5. Definición de elementos del entorno.

Elemento	Nombre	Valor inicial	Definición
Constante	NUM_EMPLEADOS	20	Cantidad de operarios disponibles
Constante	NUM_EQUIPOS	9	Cantidad de equipos
Constante	NUM_HORAS	24	Cantidad de horas por día
Constante	NUM_DIAS	92	Cantidad de días en el horizonte de planeación
Constante	CRITERIO_PARADA	2207	Número de registros del df 'REQ_PERSONAL' a recorrer
Constante	HORAS_TURNO	10	Cantidad de horas que debe ser asignado un operario de manera continua (duración de un turno normal)
Constante	HORAS_DESCANSO	14	Cantidad de horas que debe descansar un operario de manera continua luego de terminar un turno
Constante	REQ_PERSONAL	NA	Dataframe que contiene el requerimiento de personal para cada hora del día en el horizonte de planeación
Constante	COSTOS_ASIG	NA	Dataframe que contiene los costos de asignar un empleado en determinada hora del día en el horizonte de planeación
Hiperparámetro	ALPHA	0,1	Tasa de parendizaje
Hiperparámetro	GAMMA	0,8	Factor de descuento
Hiperparámetro	EPSILON	0,2	Probabilidad de exploración
Hiperparámetro	EPISODIOS	1200	Número de episodios de entrenamiento
Hiperparámetro	PENALIZACION_GRAVE	-1000000	Valor de recompensa con el que se penaliza al agente por cometer un incumplimiento grave de una política
Hiperparámetro	PENALIZACION_MODERADA	-20000	Valor de recompensa con el que se penaliza al agente por cometer un incumplimiento moderado de una política

Fuente. Elaboración propia.

Los elementos que se definen como constante en la Tabla 5 corresponde a aquellos que permanecen fijos en sus valores y estructuras durante la ejecución de todo el algoritmo. Los hiperparámetros son variables de configuración externa que no se optimizan durante el entrenamiento, sino que se definen a juicio y experiencia del investigador y cumplen la función de determinar la manera en que aprende el agente. Las variables globales del entorno corresponden a aquellas que se ejecutan de manera general en todo el entorno de aprendizaje; dentro de ellas se destacan las listas de gestión de personal, las cuales son listas temporales que van almacenando empleados disponibles, asignados y en descanso en cada momento de asignación (cada que el agente tome una acción).
En cuanto a restricciones o políticas del que deberá tener en cuenta el agente para realizar la correcta asignación de operaros a turnos serán las siguientes:

- Un operario debe trabajar turnos continuos entre 10 y 12 horas. Un turno normal tiene una duración de 10 horas, esto quiere decir que inmediatamente después de este, un operario sólo puede laborar como máximo 2 horas extra (dentro del turno de 10 horas están contenidas 2 horas de descanso).
- Un operario de cumplir con 6 turnos a la semana.
- Después de que un operario haya terminado un turno debe tener una jornada de descanso de como mínimo 14 horas antes de volver a ser asignado a un nuevo turno.
- Después de 6 turnos consecutivos de trabajo, un operario debe descansar mínimo 27 horas antes de volver a ser asignado a uno nuevo (esto para garantizar un día completo de descanso a la semana)
- Un operario sólo puede laborar como máximo 12 horas extras a la semana.
- En todo momento debe haber como mínimo un operario en toda la EDS.
- Para cada hora se debe cumplir con el requerimiento de personal calculado en la fase 1 (esta política es flexible en cuanto a su cumplimiento)

Adicionalmente en este primer paso se define la función $Reiniciar()$ la cual reinicia el entorno y hace que sus variables globales adquieran el valor inicial para ejecutarlo desde el principio. Los valores iniciales vienen dados en la cuarta columna de la Tabla 5.

5.2.2. Definición del espacio de estados y acciones.

Los estados de asignación son las situaciones en las que se encuentra el sistema (agente) en un momento dado. Para el problema dichos estados corresponden a cada una de las celdas de la matriz binaria de asignación, que es una matriz de dimensión 2208x20 (92 días x 24 horas/día x 20 empleados)

Las acciones del agente corresponden al valor que le asigna el agente a cada uno de los estados, que para el problema en teoría es decidir si la celda i, j tomará el valor de 0 o 1. Dada la complejidad del problema, se definen 3 acciones en el entorno programado de Python: asignar un turno completo de 10 horas a un empleado (0), asignar una hora extra a un empleado (1) y asignar 2 horas extras a un empleado (2). Hay otras acciones secundarias como enviar un empleado a descanso o simplemente no asignarlo, pero estas se ejecutan automáticamente dependiendo de las 3 anteriores.

5.2.3. Definición de las recompensas (función de recompensas)

Como en el proceso de asignar empleados a turnos se busca minimizar los costos totales de asignación, las recompensas del agente serán negativas, por lo que este buscará minimizar el valor de la recompensa total en cada episodio de entrenamiento. El valor de las recompensas corresponde al costo de asignar un empleado en una hora específica del día con base a un salario mínimo. Dicho costeo variará dependiendo si la hora es una hora normal o una hora extra.

Por otro lado, se tienen 2 tipos de penalizaciones (2 valores especiales de recompensa) cuando el agente incumpla alguna de las restricciones o políticas del entorno: una penalización leve que se da cuando en una hora dada el agente no asigne el suficiente personal según el requerimiento, en cuyo caso se penalizará proporcional a la cantidad de operarios faltantes; y las penalizaciones graves se da cuando el agente se quede sin operarios para asignar y una hora determinada quede sin ningún operario asignado.

Posteriormente se crea la función $paso(ACCION)$. Esta función es la más importante dentro de todo el algoritmo, pues es la que realiza la ejecución de nuestro entorno de acuerdo con la acción escogida por el agente, es decir, es la que en últimas crea el calendario de turnos para todos los trabajadores en cada episodio de aprendizaje. Esta función retornará el estado actual en el que se encuentra el entorno tras el paso de ejecución y la recompensa asociada a este paso. La forma en que penaliza esta función se esquematiza en la Figura 21.

Es de esperarse que a medida que el agente aprenda, la mayoría de las acciones tomen la ruta en color verde de la Figura 18.

5.2.4. Definición de la tabla Q

La tabla Q, también conocida como la función de valor-Q, es un componente fundamental en el algoritmo Q-learning. Esta tabla almacena los valores de calidad (Q-values) asociados a pares de estados y acciones en un entorno específico. La función de valor-Q denota la calidad de tomar una acción particular en un estado

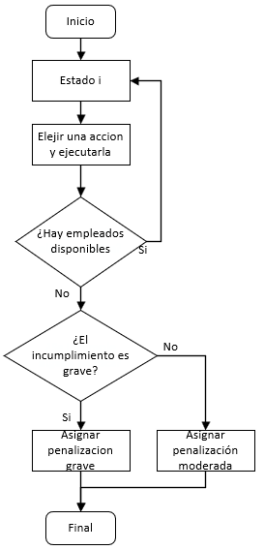

Figura 21. Diagrama de flujo para asignar recompensas a cada acción **Fuente**. Elaboración propia.

determinado y es esencial para que el agente tome decisiones informadas.

La tabla Q se actualiza durante el proceso de entrenamiento del algoritmo Q-learning. Inicialmente, los valores en la tabla Q son generalmente establecidos de manera arbitraria o a cero. A medida que el agente interactúa con el entorno, los valores se ajustan según la retroalimentación que recibe del entorno en forma de recompensas. La actualización de la tabla Q se realiza mediante la fórmula (7).

Para el problema de asignación de turnos la matriz Q se define en el entorno de Python como

Q_TABLE = np.random.uniform(low= -1, high= 1, size= [NUM_DIAS, NUM_HORAS, NUM_EQUIPOS+1, 3])

esta matriz va a ir almacenando los respectivos valores de acción para cada estado, por ello se debe mapear todos los posibles estados y a cada uno de ellos asignarle 3 espacios adicionales, para las 3 posibles acciones que se pueden realizan en el entorno. NUM_EQUIPOS se usó para definir la cantidad de empleados requeridos en un momento dado, son 10 porque el máximo puede ser 9 y el min 0 (no hay requerimiento).

5.2.5. Ejecución y entrenamiento del algoritmo Q-learning

Para iniciar el entrenamiento del algoritmo para que asigne operarios a turnos, se definen otros elementos auxiliares los cuales se muestran en la Tabla 6.

La forma en que finalmente se ejecuta el algoritmo q-learning se visualiza en la Figura 22. Cada episodio de aprendizaje o entrenamiento corresponde a una asignación completa de los operarios en el horizonte de planeación de 3 meses. En total se ejecutan 1200 episodios, y en cada uno de ellos se calcula el costo total de asignación, almacenado en la variable $RECOMPENSA_TOTAL$, y se obtiene un calendario de turnos. Se destaca que, para cada episodio de entrenamiento, si la recompensa total no es mejor que la anterior, la lista $LISTA_MEJORES_RESULTADOS$ no se actualiza.

El valor de E en el diagrama de la Figura 21 corresponde al hiperparámetro $EPSILON$ el cual se establece en 0,2 (ver Tabla 3), lo que indica que, de todas las acciones tomadas por el agente, el 20% de ellas serán aleatorias (se exploran nuevas acciones), la proporción restante se tomará de acuerdo con la que tenga mayor probabilidad de éxito según la Q_TABLE (explotar lo aprendido).

Tabla 6. Elementos para entrenamiento del algoritmo

Elemento	Nombre	Tipo	Valor inicial	Definición
Elemento de Q-Learning	NUEVO_ESTADO	Tupla	3	Variable que recibe el nuevo estado tras un paso de ejecución del algoritmo
Elemento de Q-Learning	Q_TABLE	Matriz	66240	Matriz encargada de almacenar los pesos tras el proceso de entrenamiento q-learning
Elemento de Q-Learning	LISTA_MEJORES _RESULTADOS	Lista	5	Lista encargada de almacenar los mejores resultados del entrenamiento
Elemento de Q-Learning	ESTADO	Tupla	3	Variable que almacena el estado actual antes de que el algoritmo ejecute una acción
Elemento de Q-Learning	FINAL	Booleano	FALSE	Variable que detiene la simulación

Elemento de Q-Learning	RECOMPENSA_TOTAL	Float	0	Variable que almacena el costo total después de cada episodio (similar a costo_acumulado)
Elemento de Q-Learning	ACCION	Entero	0	Variable que toma valores de 0 a 2 para definir qué tipo de turno asignar (10h, 11h 12h)
Elemento de Q-Learning	STEP	Entero	0	Variable acumulativa que cuenta los pasos necesarios para completar la asignación en cada episodio

Fuente. Elaboración propia.

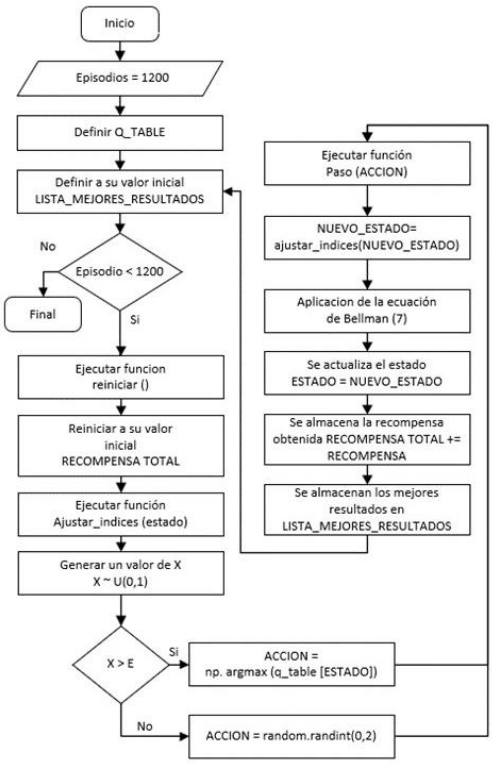

Figura 22. Diagrama de flujo para entrenamiento del algoritmo Q-Learning. **Fuente**. Elaboración propia.

5.2.6. Análisis de resultados y validación del modelo.

El primer resultado que se obtiene es el formato de calendario que se muestra en la Figura 23. El formato corresponde a un matriz binaria en donde se indica si para la hora i debe ser asignado el empleado j. Por ejemplo, se interpreta que los empleados 1 y 2 ambos inician turno de 10 horas a las 00:00 horas del día 01/07/2023.

El empleado 3 inicia su jornada a las 5:00 am y el empleado 4 a las 6:00 am, ambos con turnos normales de 10 horas para el mismo día. Para este primer calendario se obtuvo un costo total de asignación de $57.181.327. Una particularidad que se observó en este primer calendario es que las horas totales laboradas estaban muy por debajo de lo normal (240 al mes), por lo que se pensó

Día	Hora	Empleados																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	3	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	4	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	5	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	6	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	7	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	8	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	9	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	10	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	11	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	12	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	13	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
1	14	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
1	15	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
1	16	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
1	17	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
1	18	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0
1	19	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0
1	20	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0
1	21	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0
1	22	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0
1	23	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0
1	24	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0

Figura 23. Formato de calendario de turnos de personal para el día 01/07/2023 **Fuente**. Elaboración propia.

que el algoritmo usaba todos los empleados disponibles para no incurrir en penalizaciones. Dado este aspecto, se procede a correr varias veces el algoritmo disminuyendo la cantidad de empleados en cada una, obteniendo los resultados de la Tabla 7.

De la Tabla 7, como era de esperarse, la cantidad de empleados asignados es proporcional al costo total de asignación, pero inversamente proporcional a la cantidad total de horas asignadas y la cantidad de penalizaciones obtenidas. Estas cifras son totales para los tres meses, al llevarlas a promedio mensual se obtienen los datos de la Tabla 8, las cuales se pueden interpretar más fácilmente.

Recordemos que la base de horas para pagar un salario en Colombia en el año 2023 es de 240 horas al mes. Si nos fijamos en la Tabla 6, esta cifra se cumple a partir del escenario 5 que corresponde a 16 empleados. A pesar de que los costos siempre disminuyen, no hay que dejar de lado el indicador de penalizaciones, que para el escenario 5 implicaría entre 1 y 2 penalizaciones moderadas por día. A partir de este punto la decisión de la cantidad final que se deben asignar en el horizonte queda en manos de los administradores de la EDS, pues con base en su

Tabla 7. Comparación de métricas con distintas cantidades de empleados iniciales

Escenario	Cantidad de empleados	Costo total de asignación	Mínimo horas totales por empleado	Máximo horas totales por empleado	Promedio horas totales por empleado	Cantidad de penalizaciones moderadas	Cantidad de penalizaciones graves
1	20	$ 57.181.327	417	728	610	1	0
2	19	$ 57.078.975	590	725	664	10	0
3	18	$ 56.877.524	633	744	681	31	0
4	17	$ 56.611.536	662	769	703	91	0
5	16	$ 56.353.548	680	777	746	116	0
6	15	$ 55.888.140	728	792	762	229	0
7	14	$ 55.195.198	774	815	786	323	1

Fuente. Elaboración propia.

Tabla 8. Comparación de métricas con distintas cantidades de empleados iniciales (cifras promedio mensual)

Escenario	Cantidad de empleados	Costo total de asignación	Mínimo horas totales por empleado	Máximo horas totales por empleado	Promedio horas totales por empleado	Cantidad de penalizaciones moderadas	Cantidad de penalizaciones graves
1	20	$19.060.442	139,00	242,67	203,33	0,33	0,00
2	19	$19.026.325	196,67	241,67	221,33	3,33	0,00
3	18	$18.959.175	211,00	248,00	227,00	10,33	0,00
4	17	$18.870.512	220,67	256,33	234,33	30,33	0,00
5	16	$18.784.516	226,67	259,00	248,67	38,67	0,00
6	15	$18.629.380	242,67	264,00	254,00	76,33	0,00
7	14	$18.398.399	258,00	271,67	262,00	107,67	0,33

Fuente. Elaboración propia.

experiencia sopesarían los costos de asignación vs la cantidad de incumplimiento de políticas.

Como dato comparativo final, según cifras de nómina suministradas por la EDS Transpiedecuesta, para el tercer trimestre de 2023 la cantidad de promotores de servicio estuvo entre los 20 y 24 empleados al mes, que comparado con los escenarios 4, 5 y 6 (los que mejor se recomiendan) se podría obtener una reducción de hasta el 33,33% en costos.

Finalmente se evalúa el rendimiento del algoritmo en la Figura 24. Al final del entrenamiento se compararon los costos totales de asignación de cada episodio vs el costo total mínimo alcanzado en todo el entrenamiento. La tendencia decreciente de estos puntos muestra que efectivamente el agente obtenía mejores recompensas a medida que aumentaban los episodios de aprendizaje, lo que quiere decir que el aprendizaje si fue efectivo.

6. COMPARACIÓN CON MÉTODOS CONVENCIONALES DE SOLUCIÓN

Un problema convencional de asignación de turnos de personal se plantea con el siguiente modelo matemático
Parámetros de entrada:

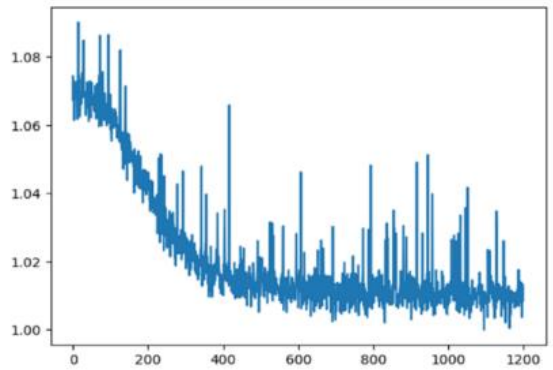

Figura 24. Costos totales relativos en cada episodio de aprendizaje **Fuente**. Elaboración propia.

- Número de turnos (n)
- Hora del día (T)
- Número de trabajadores requeridos por hora (d_t)
- Salario por turno (w_j)
- Variables de decisión

Número de trabajadores necesarios por turno de trabajo (y_j)
$a_{jt} = 1$ si el turno j cubre la ventana temporal t (j =1,..,n; t=1,..., T) y $a_{jt} = 0$ en caso contrario

Objetivo

- Minimizar el costo total de salarios pagados a todos los trabajadores

$$FO: Min\ Z = \sum_j w_j * y_j$$

Restricciones:

- Demanda por hora t que debe ser cubierta

$$\sum_j a_{jt} * y_j \geq d_t \quad t = 1, \dots, T$$
$$y_j \geq 0\ y\ entero \quad j = 1, \dots, n$$

El anterior modelo se adapta del libro titulado Applied Integer Programming: Modeling and Solution (Der-San Chen, Robert G. Batson, Yu Dang, 2010) el cual es una versión muy simplificada de un problema ETP donde ya se tienen establecidos unos turnos y el requerimiento de personal para cada hora del día. Adaptado al caso de estudio de esta investigación y sin contemplar asignación a puntos físicos (equipos), w_j denota los costos de asignar personal en cada hora, los cuales se almacenaron el dataframe $COSTOS_ASI$ y d_t corresponde al requerimiento de personal calculado en la Fase 1. Este problema se adapta y se soluciona con la librería de Python $pulp$. Bajo esta metodología se deben conocer de antemano los turnos junto con su duración, hora de inicio y hora final, los cuales se definen e la Tabla 9.

Tabla 9. Turnos de trabajo en la EDS Transpiedecuesta

Turno	Inicio	Fin	Duración
1	4:00 a. m.	1:00 p. m.	9
2	5:00 a. m.	2:00 p. m.	9
3	1:00 p. m.	10:00 p. m.	9
4	2:00 p. m.	11:00 p. m.	9
5	6:00 p. m.	4:00 a. m.	10

Fuente. Elaboración propia.

El costo total de asignación en este caso es de $59.238.912. Al extender los turnos de la Tabla 9 en las 2208 horas del horizonte de planeación se obtiene un total de 461 turnos en los que hay que asignar empleados. La solución del modelo me indica únicamente cuántos empleados debo asignar en cada uno de los turnos tal y como se muestra en la Figura 25.

A pesar de que esta metodología es más sencilla y reducida que la propuesta por Q_Learning, esta no me define el número total de empleados que debo contratar para los tres meses y cómo realizar la rotación de estos para que se cumplan sus jornadas laborales, además de que no permitiría realizar una posterior asignación a los puntos físicos de despacho (equipos). Otra desventaja que presenta esta metodología es que se deben tener establecidos turnos fijos de trabajo, además de conocerse el requerimiento de personal para cada hora, que en la mayoría de los casos reales estos parámetros se definen de forma empírica y basados en experiencia.

```
The number of workers needed for shift 1 is 2 workers
The number of workers needed for shift 2 is 3 workers
The number of workers needed for shift 3 is 3 workers
The number of workers needed for shift 4 is 7 workers
The number of workers needed for shift 5 is 0 workers
The number of workers needed for shift 6 is 2 workers
The number of workers needed for shift 7 is 2 workers
The number of workers needed for shift 8 is 3 workers
The number of workers needed for shift 9 is 5 workers
The number of workers needed for shift 10 is 0 workers
The number of workers needed for shift 11 is 2 workers
The number of workers needed for shift 12 is 2 workers
The number of workers needed for shift 13 is 2 workers
The number of workers needed for shift 14 is 0 workers
The number of workers needed for shift 15 is 4 workers
The number of workers needed for shift 16 is 2 workers
The number of workers needed for shift 17 is 3 workers
The number of workers needed for shift 18 is 3 workers
The number of workers needed for shift 19 is 7 workers
The number of workers needed for shift 20 is 0 workers
```

Figura 25. Fragmento de la solución del modelo matemático en Python **Fuente**. Elaboración propia.

Finalmente se aclara que los modelos de programación lineal, pese a su sencillez, se pueden robustecer agregando más variables y restricciones de forma que si se adapte a las necesidades de un problema como el de la EDS Transpiedecuesta, teniendo en cuenta que esto implica una mayor complejidad y métodos de solución más elaborados. No se optó por esta metodología puesto que se enmarcó el problema dentro de un contexto de analítica de datos y aprendizaje automático, además de surgió esta necesidad a raíz de la poca información en la revisión de literatura.

7. PROPUESTA DE DESARROLLO: ASIGNACIÓN DE OPERARIOS A EQUIPOS

Recordemos que el desarrollo completo del caso implica realizar una asignación de empleados tanto a turnos de trabajo como a equipos, pero dado el alcance y complejidad de esta investigación, no se desarrolla esta segunda asignación. Sin embargo, se propone continuar con este desarrollo bajo la siguiente metodología:

Dado que ya se estimó la cantidad de combustible que se demanda en cada hora y en cada equipo, y además de que ya se obtuvo un calendario de turnos de trabajo, ahora en una última fase se procede a hacer la asignación a los puntos físicos de despacho. La metodología propuesta para esta fase permite que los operarios asignados en la fase 2 tengan igualdad de condiciones en cuanto a comisiones por venta y carga laboral. Para esta fase, la asignación se debe realizar en orden cronológico ya que

a medida que transcurren las horas, los equipos asignados irán cambiando de operario.

A modo de ejemplo, vamos a suponer que para el día 01/07/2023 en la hora de las 00:00 hasta las 00:59 am la demanda de combustible en cada equipo es de d_i $i = 1,2,\ldots,9$. De la Figura 23 se sabe que están asignados los Operarios 1 y 2. En este punto se deben probar todas las posibles combinaciones de asignaciones de equipos a estos 2 operarios de forma que se minimice la diferencia en la demanda total de combustible que le corresponde a cada uno. Esta asignación permanece constante hasta la hora de las 4:00 am, pues en esta empieza turno el operario 3. Justo en este momento se le debe disminuir la carga laboral a los operarios 1 y 2 para distribuirla ahora entre 3 de forma que se garantice nuevamente la homogeneidad en demanda atendida, es decir, volver a reasignar los 9 equipos entre 3 operarios probando todas las combinaciones posibles. En este sentido un operario puede estar asignado a más de un equipo en un momento dado, pero cada uno de los 9 equipos sólo puede tener un operario asignado.

Por último, se debe tener en cuenta que los equipos surten distintos tipos de combustible y que estos tienen diferente precio de venta, por lo que implica valores de comisiones distintos. Todo lo anterior podría programarse en un algoritmo iterativo.

8. CONCLUSIONES

El resultado principal de esta metodología, a diferencia de los problemas convencionales ETP, es el nivel de detalle de los cronogramas de trabajo obtenidos. Independientemente del escenario (cantidad de empleados contratados) que se decida implementar en la empresa, el agente de aprendizaje crea dicho calendario el cual permite visualizar y gestionar a detalle la dinámica laboral de cada empleado; en este sentido al agente programa para el horizonte de planeación los turnos, horas extras y jornadas de descanso que tendrá cada empleado de forma que se satisfaga la demanda y se cumpla con la normatividad laboral.
Actualmente la toma de decisiones en las empresas se basa en datos y estadísticas, además de buscar el cumplimiento de varios objetivos a la vez. La metodología secuencial propuesta para solucionar este problema provee una gran herramienta para la toma de decisiones, pues modela datos reales de la operación y con estos se generan escenarios en los que se satisfacen las políticas empresariales y se cumplen objetivos que benefician tanto a la empresa (reducción de costos) como a los propios empleados (distribución equitativa de la carga laboral y las comisiones por ventas).
En cuanto al impacto que tiene este estudio en la empresa en la que se desarrolla el caso, se puede destacar

- La implicación en cuanto a costos, pues se puede cumplir con la operación con menos empleados de los que actualmente se tienen contratados (20 media mensual).
- Toma de decisiones con base a datos reales y propios de la operación de la empresa. Se tiene mayor control de la operación.
- Seguimiento y gestión de las jornadas de trabajo y de descanso de los empleados.
- Proyecciones presupuestales.
- Igualdad de condiciones laborales para los empleados.
- Disminución de tiempos ociosos gracias a la rotación constante del personal en puntos de despacho.

9. RECOMENDACIONES

- En cuanto a la metodología propuesta se debe realizar un análisis de variación de parámetros, para evaluar cómo influyen estos en el rendimiento del algoritmo y en los resultados.
- Se propone estudiar el problema definiendo turnos fijos de trabajo y evaluar los ya existentes.
- Dado que es un caso que se enmarca dentro del análisis de datos, se tiene la materia prima para abordar el problema a través de una simulación, empleando modelos de cola de espera con múltiples agentes.

10. ANEXOS

En el siguiente link se adjunta el código de programación en lenguaje Python

https://github.com/Michmor7/Q-Learning-for-ETP

REFERENCIAS BIBLIOGRÁFICAS

[1] J, Diaz "La retención de talento, uno de los grandes retos para las empresas en 2022". Negocios y Emprendimiento. 2022 [En línea]. Disponible en: https://www.negociosyemprendimiento.org/2022/01/retencion-talento-retos-empresas.html
[2] Bagnato, 2017
[3] Watkins y Dayan, 1992, p. 1
[4] R. S. Sutton, & A. G. Barto "Reinforcement Learning" Second Edition: An Introduction (2nd ed.). Bradford Book. 2018
[5] FS Melo, 2001, p. 3.
[6] H. Acevedo. "Metodología de asignación de turnos trabajo basado en un mecanismo de subasta aplicando al caso del servicio de pacientes en un hospital". Uniandes. 2012 [En línea]. Disponible en:

https://repositorio.uniandes.edu.co/handle/1992/24427

[7] Baeldung. "P, NP, NP-Complete and NP-Hard Problems in Computer Science". Baeldung on Computer Science. 2021 [En línea]. Disponible en: https://www.baeldung.com/cs/p-np-np-complete-np-hard

[8] A. Cortés Teoría de la complejidad computacional y teoría de la computabilidad. RISI, 1(1), 102-105. 2004

[9] S. Demassey, G. Pesant, LM." Rousseau. "Constraint Programming Based Column Generation for Employee Timetabling. In: Barták R., Milano M. (eds) Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems". CPAIOR 2005. Lecture Notes in Computer Science, vol 3524. Springer, Berlin, Heidelberg. 2005 [En línea]. Disponible en: https://doi.org/10.1007/11493853_12

[10] W. Dean "Computational Complexity Theory (Stanford Encyclopedia of Philosophy)". Stanford Encyclopedia of Philosophy. 2016 [En línea]. Disponible en: https://plato.stanford.edu/entries/computational-complexity/

[11] , A. T. Ernst, , H. Jiang, , M. Krishnamoorthy, , B. Owens, &, D. Sier "An annotated bibliography of personnel scheduling and rostering". Annals of Operations Research, 127(1-4), 21-144. 2004 [En línea]. Disponible en: doi:10.1023/B:ANOR.0000019087.46656.e2

[12] A. T. Ernst, H. Jiang, M. Krishnamoorthy, & D. Sier "Staff scheduling and rostering: A review of applications, methods and models". European Journal of Operational Research, 153(24). [En línea]. Disponible en: 2004https://doi.org/10.1016/S0377-2217(03)00095-X

[13] U. Fayyad, G. Piatetsky-Shapiro & P. Smyth "From data mining to knowledge discovery in databases". AI magazine, 17(3), 37-37. 1996

[14] J. L. González. "Tipos de aprendizaje automático – SoldAI". Medium. 2020 [En línea]. Disponible en: https://medium.com/soldai/tipos-de-aprendizaje-autom%C3%A1tico-6413e3c615e2

[15] F. Guo, & H. Song "Research and application of data-mining technique in timetable scheduling. Paper presented at the ICCET 2010 – 2010" International Conference on Computer Engineering and Technology, Proceedings, 1 V1409-V1412. 2010 [En línea]. Disponible en: doi:10.1109/ICCET.2010.5486073

[16] F. S. Hillier, G. J. Lieberman "Introducción a la investigación de operaciones" (9.a ed.). McGraw-Hill Education. 2010

[17] ICAO. "Capítulo 3 (ver.3c) – Asignación de Posiciones. En Manual de Orientación para Planificación de los Recursos Humanos" (pp. 1–16). ICAO. 2003

[18] L. P. Kaelbling, M. L. Littman & A. W. Moore "Reinforcement learning: A survey". Journal of Artificial Intelligence Research, 4, 237-285. 1996 [En línea]. Disponible en: doi:10.1613/jair.301

[19] G. P. Matos, L. M.Albino, R. L.Saldanha, & E. M. Morgado "Solving periodic timetabling problems with SAT and machine learning". Public Transport, 13(3), 625-648. 2021 [En línea]. Disponible en: doi:10.1007/s12469-020-00244-y

[20] A. Meisels, A. Schaerf "Modelling and Solving Employee Timetabling Problems". Annals of Mathematics and Artificial Intelligence 39, 41–59 2003. [En línea]. Disponible en: https://doi.org/10.1023/A:1024460714760

[21] F. S. Melo "Convergence of Q-learning: A simple proof". Institute Of Systems and Robotics, Tech. Rep, 1-4. 2001

[22] Meneses Pico, J. E. Arias Osorio "Estudio del problema de programacion de turnos de enfermería". Universidad Industrial de Santander. Escuela de Estudios Industriales y Empresariales. Tesis. 2017.

[23] C. Molnar "Interpretable Machine Learning: A Guide For Making Black Box Models Explainable". Independently published. 2022

[24] Aprende Machine Learning. "Aprendizaje por Refuerzo". 2020 [En línea]. Disponible en: https://www.aprendemachinelearning.com/aprendizaje-por-refuerzo/

[25] Aprende Machine Learning. "10 Aplicaciones de Machine Learning". [En línea]. Disponible en: https://www.aprendemachinelearning.com/aplicaciones-del-machine-learning/

[26] A. Ng, E Shyu., A. Bagul, G.L., AI Deeplearning. & Universidad de Stanford. "Supervised Machine Learning: Regression and Classification. Coursera". 2012 [En línea]. Disponible en: https://www.coursera.org/learn/machine-learning?action=enroll

[27] C. H. Papadimitriou, & K Steiglitz. (). "Combinatorial Optimization: Algorithms and Complexity". Dover Publications. 1998

[28] A.M. Printista, M.L. Errecalde, C.I. Montoya. "Una implementación paralela del algoritmo de Q-Learning basada en un esquema de comunicación con caché". [En línea]. Disponible en: http://sedici.unlp.edu.ar/bitstream/handle/10915/23363/Documento_completo.pdf?sequence=1

[29] J. E. Quintero "Creación de cronogramas de turnos de trabajo y descanso variables mediante inteligencia artificial y blockchain. - 10596/31027".

UNAD. 2020 [En línea]. Disponible en: https://repository.unad.edu.co/handle/10596/31027

[30] Rivera Jaimes, H. Lamos Diaz, D. O. Martinez Quezada, O. S. Jaramillo Piza "Solución del problema de ruteo de vehículos capacitado (cvrp) usando el algoritmo de aprendizaje reforzado q-learning". Universidad Industrial de Santander. Escuela de Estudios Industriales y Empresariales. Tesis. 2020.

[31] Rojas Villegas, J. E., Arias Osorio, A. D., Castellanos Ávila "Solución del problema de optimización integrado de programación de salas de cirugía y turnos de enfermería aplicando técnicas de optimización matemática". Universidad Industrial de Santander. Escuela de Estudios Industriales y Empresariales. Tesis. 2020 UIS

[32] C. Rubio-Terrés & A. Echevarría. "Modelos de Markov: Una herramienta útil para el análisis farmacoeconómico". Pharmacoeconomics-Spanish Research Articles. 3. 71-78. 2006

[33] A. Ruescas "Aplicación de técnicas metaheurísticas para la asignación de turnos de trabajo. Universidad Politécnica de Valencia". 2015 [En línea]. Disponible en: https://riunet.upv.es/bitstream/handle/10251/75028/TFM%20Aplicaci%C3%B3n%20de%20t%C3%A9cnicas%20metaheur%C3%ADsticas%20para%20la%20asignaci%C3%B3n%20de%20turnos%20de%20trabajo.pdf?sequence=1

[34] H. A. Taha "Investigacion De Operaciones" (9.a ed.). Pearson Educación. 2011

[35] J. Van Den Bergh, J. Beliën, P. De Bruecker, E Demeulemeester, & L. De Boeck "Personnel scheduling: A literature review". European Journal of Operational Research, 226(3), 367-385. 2013 [En línea]. Disponible en: doi:10.1016/j.ejor.2012.11.029

[36] H. Wickham, & G. Grolemund "R for data science: import, tidy, transform, visualize, and model data". O'Reilly Media, Inc.. 2016