Atribución-NoComercial-SinDerivadas 4.0 Internacional

Esta licencia es la más restrictiva de las seis licencias principales Creative Commons, permitiendo a otras solo descargar sus obras y compartirlas con otras siempre y cuando den crédito, pero no pueden cambiarlas de forma alguna ni usarlas de forma comercial.

http://creativecommons.org/licenses/by-nc-nd/4.0

UNIVERSIDAD NACIONAL “SAN LUIS GONZAGA”

FACULTAD DE INGENIERÍA DE SISTEMAS

DIRECCIÓN DE INVESTIGACIÓN

EVALUACIÓN DE ORIGINALIDAD

CONSTANCIA

El que suscribe, deja constancia que se ha realizado el análisis con el software de verificación de similitud al documento cuyo título es:

PRONÓSTICO MEDIANTE INTELIGENCIA ARTIFICIAL DE LOS EXÁMENES MÉDICO OCUPACIONAL DEL POLICLÍNICO JUAN PABLO DE LA PROVINCIA DE NAZCA PERIODO 2019-2022

Presentado por:

- **SAUÑE DELGADO ALICIA**

BACHILLER en **PREGRADO** de la facultad de Ingeniería de Sistemas. El resultado obtenido es **(porcentaje de similitud 1%)** por el cual se otorga el calificativo de:

APROBADO, según el Reglamento de Evaluación de la Originalidad.

Se adjunta al presente el reporte de evaluación con el software de verificación de originalidad.

Ica, 10 de mayo de 2024

Dr. JAVIER ORLANDO GUTIÉRREZ FERREYRA
Director de la Unidad de Investigación
Facultad de Ingeniería de Sistemas

UNIVERSIDAD NACIONAL “SAN LUIS GONZAGA”

VICERRECTORADO DE INVESTIGACIÓN

FACULTAD INGENIERÍA DE SISTEMAS

PRONÓSTICO MEDIANTE INTELIGENCIA ARTIFICIAL DE LOS EXÁMENES MÉDICO OCUPACIONAL DEL POLICLÍNICO JUAN PABLO DE LA PROVINCIA DE NAZCA PERIODO 2019-2022

Línea de investigación: Ciencias naturales, ingeniería y tecnologías sostenibles

INFORME FINAL DE TESIS

Autor: SAUÑE DELGADO ALICIA

Ica, Perú

2023

Dedicatoria

Dedico este trabajo a mis padres por el constante apoyo a lo largo de mi carrera, que hoy me ha permitido culminar esta hermosa etapa dándome cabida a nuevos proyectos y a seguir realizándome como persona y profesional.

Agradecimientos

En primer lugar, agradezco a Dios por darme salud y las fuerzas para continuar en esta hermosa carrera.

También a mi asesor de tesis, Dr. Erwin Peña Casas, por su dedicación y sabiduría.

A mis profesores que, con sus conocimientos, su paciencia y motivación apoyaron en mi crecimiento y aprendizaje.

A mis amigos y personas especiales en mi vida, que fueron un gran soporte para mi persona con el apoyo constante que me han permitido sobrellevar estos años con mucha paciencia, amor y confianza.

INDICE

Dedicatoria		ii
Agradecimientos		iii
Índice		iv
Índice de contenidos		iv
Índice de tablas		vi
Índice de figuras		vii
RESUMEN		viii
ABSTRACT		ix
I.	Introducción	1
II.	Estrategia metodológica	8
	2.1. Estrategia	8
	2.2. Metodología	9
	2.3. Herramientas	9
	2.4. Aplicación de la metodología	12
III.	Resultados	38
	3.1. Descripción de los trabajadores evaluados	38
	3.2. Estado de la Aptitud de los trabajadores	41
	3.3. Pronóstico de la Aptitud de los trabajadores	43
IV.	Discusión	44
V.	Conclusiones	45
VI.	Recomendaciones	46
VII.	Referencias bibliográficas	47
VIII.	Anexos	50
Anexo 01: Objetivos del desarrollo sostenible de las Naciones Unidas		51
Anexo 02: Evolución de la Inteligencia Artificial		52
Anexo 03: Instrumento de recojo de datos		53
Anexo 04: Empresas de los trabajadores evaluados por sexo		59

ÍNDICE DE TABLAS

TABLA I TRABAJADORES EVALUADOS SEGÚN SEXO 38
TABLA II TRABAJADORES EVALUADOS SEGÚN ESTADO CIVIL 38
TABLA III TRABAJADORES EVALUADOS SEGÚN SEXO Y ESTADO CIVIL 39
TABLA IV TRABAJADORES EVALUADOS SEGÚN SEXO DE LAS PRINCIPALES EMPRESAS 39
TABLA V MEDIDAS DE TENDENCIA CENTRAL DE CARACTERISTICAS NUMERICAS 40
TABLA VI ESTADO DE LA APTITUD DE LOS TRABAJADORES 41
TABLA VII ESTADO DE LAS APTITUDES DE LOS TRABAJADORES DE LAS EMPRESAS MAS REPRESENTATIVAS 42
TABLA VII RESULTADOS DE LAS METRICAS 43

ÍNDICE DE FIGURAS

Fig. 1. Imagen de la ubicación de la región Ica 8
Fig. 2. Imagen provincias de la región Ica 8
Fig. 3. Estructura del software WEKA 10
Fig. 4. Orange Datamining 11
Fig. 5. Formato de archivos originales 13
Fig. 6. Modelo de carga y visualización de datos 13
Fig. 7. Carga de los datos 13
Fig. 8. Información de los datos 14
Fig. 9. Características vacías de la base de datos 15
Fig. 10. Selección de las características 16
Fig. 11 Base de datos filtrada 17
Fig. 12. Modelo de evaluación previa 17
Fig. 13. Eliminación de características que tienen muchos valores perdidos 18
Fig. 14. Información de la base de datos con preprocesado inicial 18
Fig. 15. Inputacion de datos 19
Fig. 16. Información de los datos después del 2da prepocesamiento 20
Fig. 17. Modelo orange para el preprocesado de los datos 20
Fig. 18. Características (Features) finales 21
Fig. 19. Carga de datos finales 22
Fig. 20. Análisis descriptivo de los datos numéricos 23
Fig. 21. Widget para el análisis de datos categoricos 24
Fig. 22. Analisis de la Aptitud según el estado civil 24
Fig. 23. Análisis de la aptitud según grupo sanguíneo 25
Fig. 24. Análisis de aptitud según consumo de alcohol 25
Fig. 25. Análisis de aptitud según hábitos de fumar 26
Fig. 26. Análisis de aptitud según consumo de drogas 26
Fig. 27. Análisis de aptitud según su grado de instrucción 27
Fig. 28. Análisis de aptitud según conclusiones oftalmológicas 27
Fig. 29. Análisis de aptitud según conclusiones del triaje 28
Fig. 30. Características relevantes para la predicción 29
Fig. 31. Transformación de las características 30
Fig. 32. Aplicación de la técnica del One hot encoding 30
Fig. 33. Transformación de características numéricas 31

Fig. 34. Otras características transformadas 32
Fig. 35. Modelo para el análisis y transformación de los datos 32
Fig. 36. Modelos de predicción para la característica AptitudFinal 33
Fig. 37. Resultado de las métricas 34
Fig. 38. Resultados con 75% de datos de entrenamiento 35
Fig. 39. Gráfica de la curva AUC 35
Fig. 40. Configuración del algoritmo de regresión logística 36
Fig. 41. Resultados de la predicción 36
Fig. 42. Resultados de la predicción 37

RESUMEN

La presente tesis de investigación tuvo como objetivo realizar el pronóstico de los exámenes médico ocupacional de los pacientes del policlínico Juan Pablo de la ciudad de Nazca periodo 2019-2022, El estudio del nivel predictivo, para un enfoque cuantitativo, del tipo aplicado, utilizó los datos registrados en el periodo seleccionado de los trabajadores de las empresas que se realizaron la evaluación para la aptitud o no para el trabajo, la metodología utilizada CRISP-DM (Cross Industry Standard Process for Data Mining) empleada para proyectos de dataming e inteligencia artificial, y como herramienta se seleccionó el software Orange dataming por su aplicación gráfica para el lenguaje python, que no requieren de conocimientos de programación. Los resultados obtenidos para diversos algoritmos de predicción arrojan una precisión de cerca del 80% para la regresión logística y las redes neuronales. La investigación concluye que es muy importante realizar el pronóstico en este campo, que sirve como ayuda a los profesionales de la salud en el análisis del diagnóstico la aptitud de los trabajadores de las empresas que hacen uso del policlínico Juan pablo.

Palabras clave: Aptitud, algoritmos, Inteligencia artificial, exámenes médicos, Orange Datamining.

ABSTRACT

The present research thesis aimed to perform the prognosis of occupational medical examinations of patients of the Juan Pablo polyclinic of the city of Nazca period 2019-2022, The study of the predictive level, for a quantitative approach, of the applied type, used the data recorded in the selected period of the workers of the companies that were performed the evaluation for fitness or not for work, The methodology used was CRISP-DM (Cross Industry Standard Process for Data Mining) used for dataming and artificial intelligence projects, and the Orange dataming software was selected as a tool because of its graphical application for the Python language, which does not require programming knowledge. The results obtained for various prediction algorithms show an accuracy of about 80% for logistic regression and neural networks. The research concludes that it is very important to perform forecasting in this field, which serves as an aid to health professionals in the analysis of the diagnosis of the aptitude of the workers of the companies that make use of the Juan Pablo polyclinic.

Keywords: Fitness, algorithms, artificial intelligence, medical examinations, Orange Datamining.

I. INTRODUCCIÓN

Muchos estudios se están realizando en el campo de la investigación en el mundo, en donde la inteligencia artificial esta aportando enormemente en el campo de la medicina, estudios como el de [1]–[3] para el caso de las defunciones, similarmente estudios sobre los clasificadores de Inteligencia artificial en el campo de la medicina [3]. Como se desprende de estos estudios el incremento de la tecnología, la innovación en el campo de la medicina es un buen indicador de que se va por buen camino, todo relacionado con los "La Agenda 2030 y los Objetivos de Desarrollo Sostenible" de las Naciones Unidas, específicamente en el campo de Objetivo 03 Salud y Bienestar [4].

Aspectos científicos

Investigación donde se analiza, el crecimiento de los datos masivos de las bases de datos, además de los diversos parámetros en el campo de los resultados médicos, el Machine Learning o aprendizaje automático de la Inteligencia artificial como herramienta encargada para procesar estos datos y encontrar patrones para el diagnóstico de pacientes con angina de pecho de las anotaciones clínicas; al mismo tiempo predecir la mortalidad por este tipo de enfermedad en la hemorragia cerebral, como también detectar pacientes con insuficiencia cardíaca de los registros de las atenciones médicas en las bases de datos [5].

En el proyecto de titulación, la investigación tuvo como objetivo implementar un módulo predictivo por medio los algoritmos de entrenamiento supervisado, con herramientas de inteligencia artificial (Machine Learning/ML), en este caso para el caso del área médica en la que este tipo de avance tecnológico de los servicios en la nube o web cuentas con una mayor eficacia para el diagnóstico en la medicina, como apoyo de los médicos. Con tal los médicos especialistas de la medicina tendrán a su disposición estas herramientas como apoyo a la telemedicina, e innovación en el trabajo y servicios de salud. La investigación analiza diversas enfermedades que se pueden presentar en un paciente, y con algoritmos de clasificación y codificación aplicando codificación

CIE-10 para una diversidad de características o variables. Con relación a la implementación del algoritmo, se hace uso de técnicas de minería de datos, y algoritmos para procesamiento del lenguaje natural, con registros numéricos y de textos plano del hospital de Cayapas. Para la predicción de los diagnósticos se utilizaron algoritmos de árboles de decisiones para realizar el agrupamiento y clasificación de los patrones que permitan calcular el valor requerido y tener una buena decisión en el pronóstico. Los patrones resultantes fueron etiquetados por medio de cálculos matemáticos obteniendo grupos con características similares [6].

En la investigación relacionada con las tecnologías innovadoras en el campo de las ciencias médicas, en las que se están realizando una de las mayores aplicaciones para la predicción de tratamientos y diagnósticos médicos, mejorando y optimizando los diagnósticos en este campo. Esta tecnología informática está ayudando a encontrar enfermedades en etapas tempranas. El campo de estas tecnologías como la Inteligencia artificial aplicado a las ciencias médicas, orientado a reducir los problemas que se dan en los hospitales de Lima, con esta investigación se propone la implementación de un chatbot con Inteligencia Artificial para el diagnóstico de enfermedades por su enorme capacidad de diagnóstico, el estudio cualitativo identifica las mejoras permisibles con la implementación de estas tecnologías innovadoras orientadas a la Inteligencia artificial [7].

En el mismo campo de la medicina el estudio sobre el impacto de la Inteligencia artificial (IA) en pleno auge de la pandemia desde el año 2020 en América Latina, donde de evalúa la evolución y conocimientos del uso de la AI. El estudio está orientado al monitoreo de diversas aplicaciones de la AI para el caso de la emergencia sanitaria. El estudio igualmente profundiza sobre el estudio de experiencias, así como los inconvenientes que se presentaron como también las oportunidades de estas tecnologías en la región de América Latina pasado la pandemia del COVID-19 [8].

En el caso de esta investigación en la que se analiza la información de los pacientes de consulta externa del Hospital Regional de Moquegua, para ello se realizó la revisión bibliográfica sobre investigaciones relacionadas con la metodología de Inteligencia artificial y algoritmos posibles de emplear y evaluar y seleccionar una específica. Con la evaluación de estos algoritmos se determinó la utilización del algoritmo de regresión logística con la que se permite evaluar, clasificar y hacer pronósticos de resultados esperados según el objetivo establecido. La emergencia sanitaria ocasionada por el COVID-19 limitó la atención de este servicio de consulta externa desde marzo 2020 hasta junio 2021, por lo que no se cuenta con información de la posible demanda del servicio por cada uno de los tópicos. En la investigación se plantea la predicción, con características como la edad, sexo, cantidad de pacientes femenino que requirieron alguna especialidad para esta consulta. El algoritmo de regresión logística utilizado de la herramienta scikit-learn permitió tener la

clasificación y determinación de los grupos más relevantes según lo planteado en el objetivo, siendo dos variables de mucho interés la edad y el sexo. El modelo con el algoritmo de regresión logística presentó resultados que no reflejaron una correspondencia real con la de la predicción. Se concluye en determinar que el modelo del algoritmo planteado requiere de una mayor cantidad de variables que debe ser incorporadas al modelo para tener una predicción más precisa [9].

Este otro estudio igualmente en el campo de la medicina, para el cáncer de mama tiene mucha relevancia para un adecuado diagnóstico y tratamiento que se debe seguir el paciente, siendo esto un enorme apoyo a la labor del médico. Según la investigación el personal médico especialista actualmente está empleando estas tecnologías con actividades que no usan herramientas para tratamiento no numérico. En este entender las redes neuronales artificiales (RNA) se presentan con una de las herramientas más poderosas para el análisis de los datos en las que estas cuentan con relaciones no lineales para una adecuada predicción. En la investigación se estima la probabilidad de clasificación mediante el algoritmo con la regla de Bayes para un modelo de perceptrón multicapa, con el que se permite tener la mejor predicción para la regla establecida. La investigación aplica la predicción de recidiva de pacientes que han sido operados de cáncer de mama. Los datos fueron obtenidos de la información clínica-patológica con características como: edad de la paciente, tamaño del tumor, receptores de estrógenos, datos obtenidos del servicio de oncología del hospital de la universidad de Málaga, la estructura del perceptrón multicapa fueron variadas por prueba y error hasta obtener el mejor modelo en su precisión. Posterior al entrenamiento del modelo y con los resultados obtenidos el modelo es adecuado para las predicciones sobre la recidiva para diversos periodos de tiempo [10].

En el mismo camino del empleo de la inteligencia artificial en la ciudad de Medellín Colombia, la investigación cuya finalidad es maximizar los recursos disponibles que permitan atender la enorme demanda que se genera en el hospital, más aún en un área como es a la unidad de cuidados intensivos (UCI) área que debe tener equipamiento de la más alta tecnología para lograr tener a los pacientes monitoreados las 24 horas, pero que resulta limitado esta unidades, por lo que obliga a encontrar nuevas alternativas para atender a más pacientes con los mismos recursos. La investigación plantea el diseño e implementación de la inteligencia artificial, como soporte a la administración de estas UCIs prediciendo el uso basado en los datos del paciente, garantizando una mejor planificación y aprovechamiento de los escasos recursos. En la investigación se entrenaron diversos modelos de AI evaluando la eficiencia de cada uno de ellos en cuanto a su nivel de predicción para los tiempos de estadio en la UCI, con ello se pudo seleccionar el más eficiente para el hospital. Los resultados de los algoritmos de regresión y de clasificación categórica donde se

rescata los árboles de decisión (Random Forest) con 69% de precisión, la Naives bayes con 64.3 % de precisión, asimismo en los modelos de regresión se destacó el XGBoost con 7.43 de desviación estándar para los días de estadio [11].

Otros estudios en el campo de la medicina, en esta investigación sobre el procesamiento de imágenes de pacientes con dolencia cardiaca, ya que esta influye de sobremanera en las actividades del trabajo clínico: interacción médico paciente, obtención de la imagen, informes del paciente, diagnósticos, pronósticos e intervenciones del médico. El conocimiento de estas imágenes por medio de la investigación clínica, la creciente incorporación de la inteligencia artificial en el campo de la cardiología, que demuestra cómo puede aportar por medio de la automatización e integración de los datos. La investigación tuvo como objetivo a revisión de cómo se incorpora la inteligencia artificial en el trabajo de atención clínica de los pacientes específicamente la imagen cardiaca, aunque desde una vista sistémica para la gestión de toda la información, que permitan encontrar las limitaciones que deben ser superadas para una implementación más general [12].

La investigación realizada en una institución prestadora de salud nivel III de Lima centro 2022 con la finalidad de determinar la influencia de la inteligencia artificial y gobierno digital en el periodo de COVID-19, la investigación de tipo cuantitativo aplicada, cuyo diseño no experimental de corte transversal y de causa efecto. Para la investigación se utilizó una población de 2860 empleados de salud, de los que se seleccionó una muestra probabilística simple de 339 empleados. Para la recolección de datos se utilizó la entrevista como técnica, con un cuestionario validado mediante la prueba del alfa de Cronbach con resultados para la inteligencia artificial de 0.953 y para el caso de gobierno digital con 0.908 utilizando para la prueba estadística el software SPSS versión 26. Se obtuvo el análisis descriptivo e inferencial, y la evaluación con el algoritmo de regresión logística ordinal, con un p valor de 0.05, y un chi cuadrado de 231.094 con un nivel de significancia de 0.000 y un valor Nagelkerke de 0.615 vale decir una influencia del 61.15% para ambos indicadores [13].

En este siguiente estudio donde se tiene que el aprendizaje automático o Machine Learning (ML) es considerada una extraordinaria herramienta utilizada en muchas industrias. En el estudio como consecuencia de la gran cantidad de información disponibles sobre exámenes clínicos que se almacenan de forma digital en el campo de la medicina ofrece un campo ideal para la aplicación de estas nuevas tecnologías, ya que este ML demuestra un enorme potencial para apoyar y mejorar los sistemas de salud debido a los millones de datos que se almacenan en el tiempo y que siguen creciendo día a día, logrando diseñar modelos para predicción. A pesar de esta potencialidad sobre todo del uso de métodos algorítmicos se puede mejorar la calidad de los sistemas de atención a

pacientes, sin embargo, es aún requerido una mayor validación de estos procesos con estas tecnologías [14].

En la investigación sobre la Radiómica del TC basal, para evaluar la clasificación de Inteligencia Artificial con la finalidad de predecir el pronóstico funcional para pacientes con HIE. Para ello se utilizó 105 biomarcadores de imágenes cuantitativa del TC basal del mismo número de pacientes HIE. Los marcadores se obtuvieron del software Radiomics de 3D-Slicer que fueron segmentados manualmente con un radiólogo especialista. El procesamiento de los datos requirió realizar la imputación de los datos perdidos y valores atípicos, los mismos que posteriormente fueron estandarizados y normalizados. Para las pruebas los datos fueron divididos en 70%- 30% para el entrenamiento y validación respectivamente. Los métodos de selección de variables requirieron clasificadores predictivos lineales, no lineales y ambos la clasificación del pronóstico funcional para la clasificación buen/mal (mRS0- 2/mRS3-6). Para el entrenamiento se realizó 10 iteraciones con validación cruzada, se calculó la media de dichos valores para el indicador AUC (área bajo la curva). Posteriormente se realizó el cálculo de la sensibilidad de cada cohorte de validación. La clasificación mediante ANOVA y P-SVM (soporte de máquina vectorial), KNN (vecinos más cercanos) y DF-10, estos arrojaron una AUC de 0.798, 0.752 y 0.742 correspondiente a cada algoritmo. En cuanto al modelo con mejor rendimiento la sensibilidad fue de 0.892 y con 0 falsos negativos. Con estos resultados se concluye que el empleo de clasificadores con AI para Radiómica tienen un nivel de predicción muy prometedor para el pronóstico funcional para este tipo de pacientes, sin duda hay que mejorar el rendimiento de los algoritmos y validarlos con cohortes externos [3].

El siguiente proyecto para la aplicación de la transformación digital para pacientes que tengan alguna enfermedad o necesidad de diagnóstico médico. En la investigación se persigue encontrar una alternativa que acoja la problemática utilizando las TICs. Se utiliza la plataforma de Re-Core que permite que los pacientes tener su historial médico de manera integrada, con lo cual se tiene una mejor administración, con mayor seguridad, eficiente y eficaz. Los pacientes pueden registrar la información para que los médicos que los atiende tengan el control sobre los medicamentos, así como sus evaluaciones y visitas. Aquí se destaca las sugerencias del algoritmo de inteligencia artificial para tener una mejor comprensión sobre la información de estos pacientes y poder actuar de manera rápida ante alguna dolencia [15].

Igualmente, en el mismo camino en los servicios médicos la investigación realizada en la empresa MOK del Perú en relación con sus servicios de asistencia médica, para personas particulares como empresas. La investigación tuvo como objetivo desarrollar una propuesta para mejorar el proceso de

sus servicios médicos como ke de consultas médicas, atención domiciliaria y servicio de ambulancias, esta propuesta basada en la inteligencia artificial tiene como fin mejorar sus servicios haciendo más eficiente estos procesos los resultados han conseguido identificar que los servicios generan el 43% de lo ingresos de la empresa MOK Perú. Se detectó los errores en el proceso de atención a clientes del Call Center relacionados con las llamadas para solicitar los servicios de la empresa en la que el 34% de estas llamadas no son atendidas. La investigación evaluó la situación actual del servicio y la situación futura con la implementación de un chatbot, para ello se requiere de una inversión de 255,000 nuevos soles para un crecimiento de la rentabilidad del 5% en el año 5, además de conseguir los indicadores de gestión como son la disminución del tiempo de espera de 22 a 2 segundos, el tiempo promedio de operación de 11.8 minutos a 7.5 minutos y las limitaciones en cuanto a las llamadas no atendidas reducirlas del 34% al 16% [16].

Como todos los estudios e investigaciones anteriores, en la presente investigación orientada a conocer los principales algoritmos, técnicas utilizadas de la inteligencia artificial para la lucha contra la COVID-19. En esta lucha el diagnóstico y desarrollo de nuevas vacunas para el pronóstico de COVID-19. En la investigación se realizó la búsqueda bibliográfica en diversas bases de datos sobre el tema y tener una descripción sobre las aplicaciones innovadoras en este campo de la ciencia hasta el 2020. Se han desarrollado diversos algoritmos con relación al COVID-19 para pronosticar la detección de problemas pulmonares de las imágenes de rayos X de los pacientes. Aquí los algoritmos de Deep Learning como el caso de las CNN (Convolucional Neural Network/Redes neuronales convolucionales) han arrojado ser las mejores predictoras para este tipo de soluciones, para pronosticar si un paciente tiene COVID-19 sí o no [17].

Realidad problemática

Desde el 2018 el policlínico Juan Pablo de la provincia de Nazca, viene realizando la atención de los pacientes de diversas instituciones para determinar la aptitud o no de dichos pacientes, esta evaluación es obligatoria, de tal manera que se determine la aptitud o no de los trabajadores para conocimiento de estas empresas y puedan ello tomar las decisiones que estimen conveniente. Desde esta fecha se ha registrado la información sobre dichas atenciones de exámenes médico ocupacional generando grandes cantidades de datos, pero que a la fecha esta continua en las bases de datos de los sistemas del policlínico y no se le está aprovechando en tener conocimiento de todos esos datos. Existe muchos datos, pero poco conocimiento. Ante esta realidad como encargada del manejo y administración de la información de policlínico en mi calidad de bachiller en Ingeniería de Sistemas con el conocimiento recibido en mi formación profesional, y que deseando aportar en bien la institución en donde laboro, para obtener diversos conocimientos y patrones de los datos

registrados, además de lograr mi meta de cumplir con la obtención de mi titulo profesional. Antes esto me pregunto.

¿Cuál será el nivel de pronóstico de exámenes médico ocupacional del policlínico Juan Pablo de la ciudad de Nazca periodo 2019-2022?

Justificación

La investigación no solo es importante por lo que puede aportar al policlínico, sino reviste una mayor importancia debido a que se va a utilizar herramientas tecnológicas que están relacionadas con el avance tecnológico, como es la Industria 4.0 con muchos aportes en la empresas en la que se incluye herramientas como la inteligencia artificial. Además de ello es importante por el nuevo conocimiento que generará la aplicación de la tesis en el uso de algunos algoritmos para el pronóstico de las características y patrones que pueden desprenderse de estos.

Por lo tanto, no solo se apoya a la institución en tener conocimiento sobre los exámenes médico ocupacional, sino que igualmente se aporta en el conocimiento del uso de la inteligencia artificial y sus algoritmos para el nivel de predicción de características, por un lado, y los resultados para los pacientes de las empresas que acuden a policlínico.

Con todo lo mencionado, consideramos que la investigación con el planteamiento y el desarrollo de del proyecto se justifica plenamente.

Para lograr cumplir con el desarrollo del proyecto nos planteamos el objetivo de "Realizar el pronóstico de los exámenes médico ocupacional de los pacientes del policlínico Juan Pablo de la ciudad de Nazca periodo 2019-2022". Y en la que consideramos algunos objetivos específicos que guiaron nuestro desarrollo mencionado en las líneas siguientes:

1. Describir los tipos de pacientes/empresa que se realiza exámenes médico ocupacional de los pacientes del policlínico Juan Pablo de la ciudad de nazca periodo 2019-2022
2. Evidenciar el estado de aptitud de los pacientes/empresa que se realiza exámenes médico ocupacional de los pacientes del policlínico Juan Pablo de la ciudad de nazca periodo 2019-2022
3. Pronosticar la aptitud de los exámenes médico ocupacional de los pacientes del policlínico Juan Pablo de la ciudad de nazca periodo 2019-2022

II. ESTRATEGIA METODOLÓGICA

2.1 Estrategia

La investigación que se aplica al presente proyecto de tesis es del tipo aplicado, correspondiendo al nivel de investigación predictivo. La población para el estudio corresponde a los exámenes médicos ocupacionales realizados en el periodo del 2019 al 2022, correspondiendo a un estudio retrospectivo. El estudio es de enfoque cuantitativo.

Los datos han sido obtenidos del sistema de registros de las atenciones realizadas para examen médico ocupacional del policlínico Juan Pablo de la ciudad de Nazca en la Región de Ica y que han sido extraídos en un archivo de EXCEL, cuya estructura y parte de estos datos se muestran en el Anexo 01 (instrumento de recojo de datos). esta estructura cuenta con 254 columnas de datos, muchos de ellos no cuentan con información, es decir se encuentran vacíos como se puede comprobar en el anexo 01.

Fig. 1. Imagen de la ubicación de la región Ica
Fuente: https://commons.wikimedia.org/wiki/File:Diócesis_de_Ica.svg

Fig. 2. Imagen provincias de la región Ica
Fuente: https://commons.wikimedia.org/wiki/File:División_Politica_de_Ica.jpg

2.2 Metodología

La metodología utilizada en el presente trabajo es la CRISP-DM (Cross Industry Standard Process for Data Mining) es un modelo de referencia que describe de forma general las fases, tareas generales y salidas de un proyecto de Data Mining en general. Por otro lado, la guía del usuario proporciona información a detalle sobre la aplicación del modelo a proyectos de Data Minning específicos, se proporcionan guías y listas de cotejo sobre las actividades y tareas correspondientes a cada fase. De esta manera se estructura el ciclo de vida de un proyecto en seis fases, que interactúan entre ellas de forma iterativa durante el desarrollo de este[18].

Las fases de la metodología corresponden a: 1) Comprensión del negocio, 2) Comprensión de los datos, 3) Preparación de los datos, 4) Modelado, 5) Evaluación del modelo y 6) la implementación.

2.3 Herramientas

Para cumplir con la aplicación de la metodología se evaluó dos herramientas para el tratamiento de los datos por un lado WEKA y la Otra Orange Data Mininig

Weka es una colección de algoritmos de aprendizaje automático para tareas de minería de datos. Contiene herramientas para la preparación, clasificación, regresión, agrupación en clústeres, extracción de reglas de asociación y visualización de datos. Weka es un software de código abierto emitido bajo la Licencia Pública General GNU. Dentro de esta herramienta se puede aplicar el ML (Machine Learning) o aprendizaje automático, el mismo que cuenta con un conjunto de tecnologías para analizar datos y hacer predicciones, inclusive sus últimos avances incluyen tratamiento para inteligencia artificial [19].

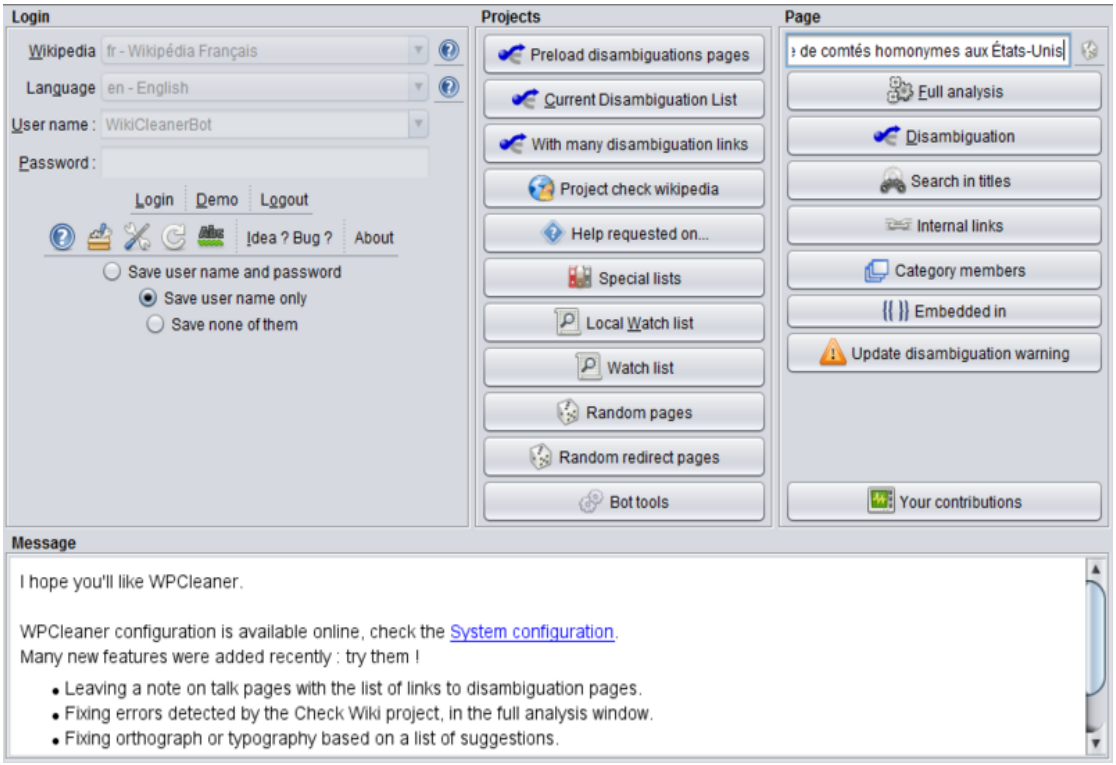


Fig. 3. Estructura del software WEKA

Fuente: https://commons.wikimedia.org/wiki/File:WPCleaner_-_Logged.png

Orange Data mining, Orange es un paquete de software de programación visual de código abierto basado en componentes para visualización de datos, aprendizaje automático, minería de datos y análisis de datos. Los componentes naranjas se llaman widgets. Van desde la simple visualización de datos, la selección de subconjuntos y el preprocesamiento hasta la evaluación empírica de algoritmos de aprendizaje y modelos predictivos [20].

Utilizado en escuelas, universidades y cursos de formación profesional en todo el mundo, Orange admite formación práctica e ilustraciones visuales de conceptos de ciencia de datos. Incluso hay widgets que fueron diseñados especialmente para la enseñanza.

Después de sólo unas pocas horas de formación en clase, comprenderá los algoritmos clave de aprendizaje automático y podrá aplicarlos a una amplia gama de problemas. Sin codificación ni matemáticas: ¡solo visualizaciones y exploración de datos interactiva! (https://orangedatamining.com/workshops/)

Los widgets de Orange son componentes básicos de los flujos de trabajo de análisis de datos que se ensamblan en el entorno de programación visual de Orange. Los widgets se agrupan en

clases según su función. Un flujo de trabajo típico puede combinar widgets para la entrada y filtrado de datos, visualización y extracción de datos predictivos (https://orangedatamining.com/getting-started/)

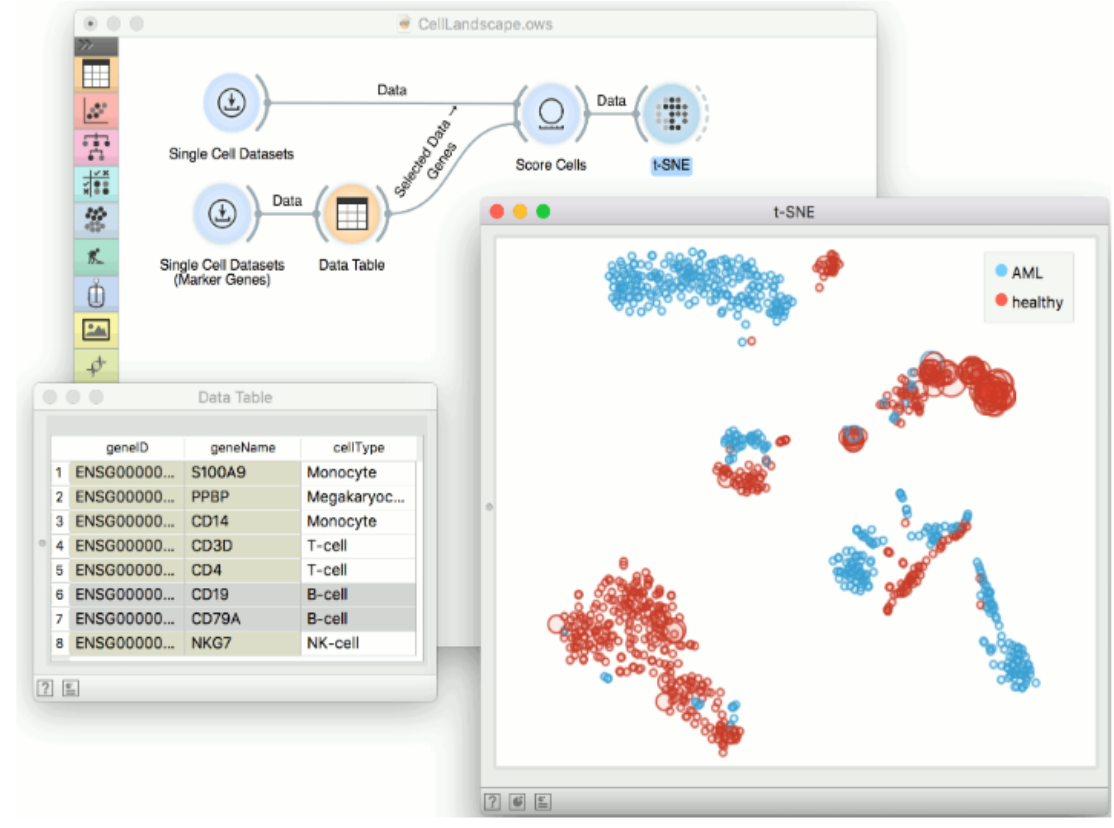


Fig. 4. Orange Datamining

Fuente: https://orangedatamining.com/getting-started/

De las dos herramientas evaluadas, se pudo determinar que ambas herramientas son fáciles de utilizar por manejo gráfico de las aplicaciones de los algoritmos y análisis estadístico de los datos. igualmente, cualquiera de las herramientas pudo ser utilizada debido a que ambas son de uso libre, no requiere licencia de uso.

Para nuestra investigación, se seleccionó que la mejor herramienta para el tratamiento de los datos y la predicción a realizar el software Orange datamining era la ideal, y es la que se aplica en toda la metodología.

2.4 Aplicación de la metodología

Fase 1: Comprensión del negocio, en esta primera fase se identifica el objetivo del negocio, cual es la problemática para lograr este objetivo y reflejarlos en un plan de proyecto.

El policlínico San Pablo, dada su actividad en el área de la salud, la misma que esta destinada a la atención de los pacientes que llegan a realizarse diversas atenciones médicas. En ese camino realiza el diagnóstico de los pacientes que pertenecen a diversas instituciones para determinar su estado relacionado con su estado médico ocupacional, en la que se tomar diversas pruebas de diagnóstico para conocer el estado de cada persona y determinar la aptitud para el trabajo desde el punto de vista de la salud.

En por ello que como institución conociendo que se registra gran cantidad de datos relacionados con este diagnóstico (ver anexo 01), es que nos trazamos el objetivo de como por medio de estos datos pronosticar el comportamiento del análisis médico ocupacional de los pacientes de las empresas, tomando diversas variables de estos pacientes, como edad, sexo, y otras variables que permitan el pronóstico de los exámenes medico ocupacionales y mejorar el servicio de esta atención. Considerando un problema que se este almacenando gran cantidad de datos de las atenciones médicas para el diagnóstico médico ocupacional, y que esta información no este siendo aprovechada, más aun en un escenario actual donde la inteligencia artificial esta proporcionando una serie de herramientas y algoritmos para aprovechar esos datos en instituciones médicas, como el caso de [2] en la aplicación de la inteligencia artificial mediante el uso de Machine Learning para el proceso de clasificación de datos asociados al centro médico CMC.

Fase 2: Comprensión de los datos, para lograr esta fase, es el primer paso para conocer el problema, aquí se acopian los datos iniciales (datos en bruto), se realiza una descripción de los datos, y se realiza el análisis exploratorio de estos datos. para ello se utilizan diversas pruebas estadísticas básicas, además de comprobar la calidad de los datos.

Para lograr un análisis de los datos primero se realizar la carga de los datos en el software Orange y a partir de allí con los widgets necesarios se realiza la comprensión de los datos. el archivo extraído de las fuentes de registro del policlínico se extrajo en formato Excel, el cual para un mejor tratamiento de los datos se convirtió a un archivo de valores o archivo CSV, que son archivos más pequeños que el archivo inicial de Excel.

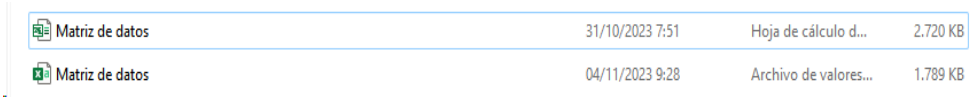


Fig. 5. Formato de archivos originales

En ambos casos los archivos tienen en mismo nombre Matriz de datos, pero se diferencia por el formato guardado, se comprueba que el archivo de valores (CSV) tiene un peso de 1789 Kb, cerca de menos 1000 kb en relación con el archivo original.

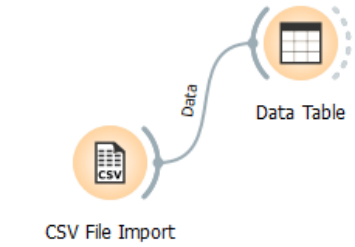
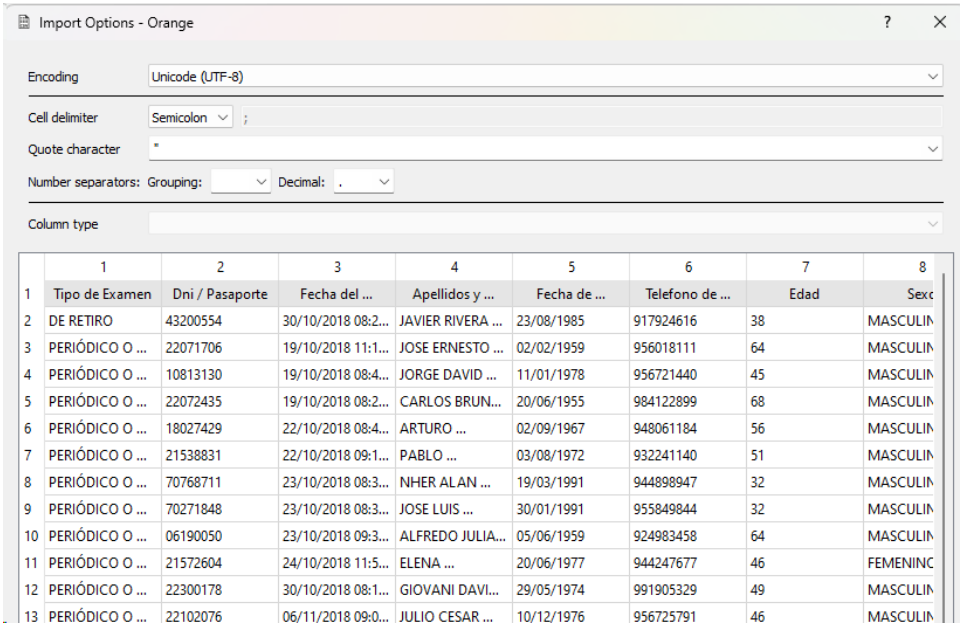


Fig. 6. Modelo de carga y visualización de datos

Los datos se cargan con el widget CSV File Import y se visualiza con el widget Data Table



	1	2	3	4	5	6	7	8
1	Tipo de Examen	Dni / Pasaporte	Fecha del ...	Apellidos y ...	Fecha de ...	Telefono de ...	Edad	Sexo
2	DE RETIRO	43200554	30/10/2018 08:2...	JAVIER RIVERA ...	23/08/1985	917924616	38	MASCULIN
3	PERIÓDICO O ...	22071706	19/10/2018 11:1...	JOSE ERNESTO ...	02/02/1959	956018111	64	MASCULIN
4	PERIÓDICO O ...	10813130	19/10/2018 08:4...	JORGE DAVID ...	11/01/1978	956721440	45	MASCULIN
5	PERIÓDICO O ...	22072435	19/10/2018 08:2...	CARLOS BRUN...	20/06/1955	984122899	68	MASCULIN
6	PERIÓDICO O ...	18027429	22/10/2018 08:4...	ARTURO ...	02/09/1967	948061184	56	MASCULIN
7	PERIÓDICO O ...	21538831	22/10/2018 09:1...	PABLO ...	03/08/1972	932241140	51	MASCULIN
8	PERIÓDICO O ...	70768711	23/10/2018 08:3...	NHER ALAN ...	19/03/1991	944898947	32	MASCULIN
9	PERIÓDICO O ...	70271848	23/10/2018 08:3...	JOSE LUIS ...	30/01/1991	955849844	32	MASCULIN
10	PERIÓDICO O ...	06190050	23/10/2018 09:3...	ALFREDO JULIA...	05/06/1959	924983458	64	MASCULIN
11	PERIÓDICO O ...	21572604	24/10/2018 11:5...	ELENA ...	20/06/1977	944247677	46	FEMENINO
12	PERIÓDICO O ...	22300178	30/10/2018 08:1...	GIOVANI DAVI...	29/05/1974	991905329	49	MASCULIN
13	PERIÓDICO O ...	22102076	06/11/2018 09:0...	JULIO CESAR ...	10/12/1976	956725791	46	MASCULIN

Fig. 7. Carga de los datos

Al cargar los datos con el widget CSV File Import, es importante considerar el delimitador de celdas (Cell delimiter), esto debido a que el archivo convertido de Excel en formato CSV, este puede tener un delimitador con coma (,) o con punto y coma (;).

Una vez cargado el archivo la primera evaluación que se realiza es sobre el contenido del archivo CSV. El mismo se presenta en la Fig. 8.

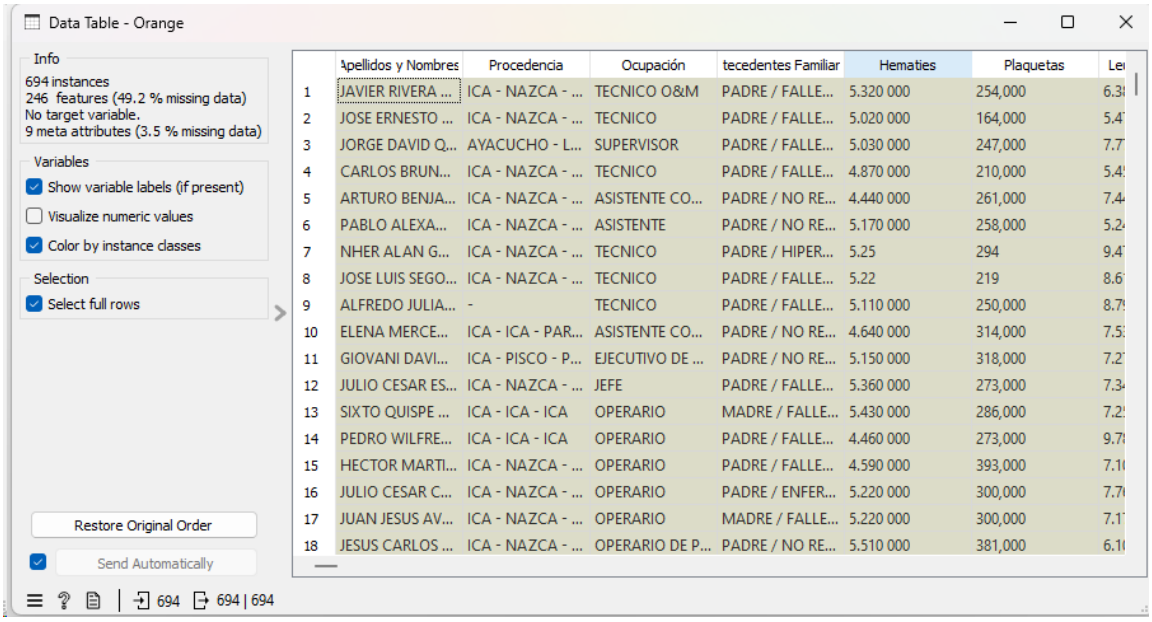


Fig. 8. Información de los datos

En la Fig. 8, los datos presentados con el widget Data Table, la información indica que existen 694 instancias (registros de datos) en el periodo seleccionado. El archivo contiene 246 características (columnas), en las que se tiene 49.2% de datos perdidos o vacíos.

En la parte derecha del widget, se puede ver el contenido de las características. Realizado el recorrido por todas las columnas o características de los datos, se puedo comprobar que existen muchas de estas características cuyos contenidos se encuentran vacíos como se aprecia en la Fig. 9.

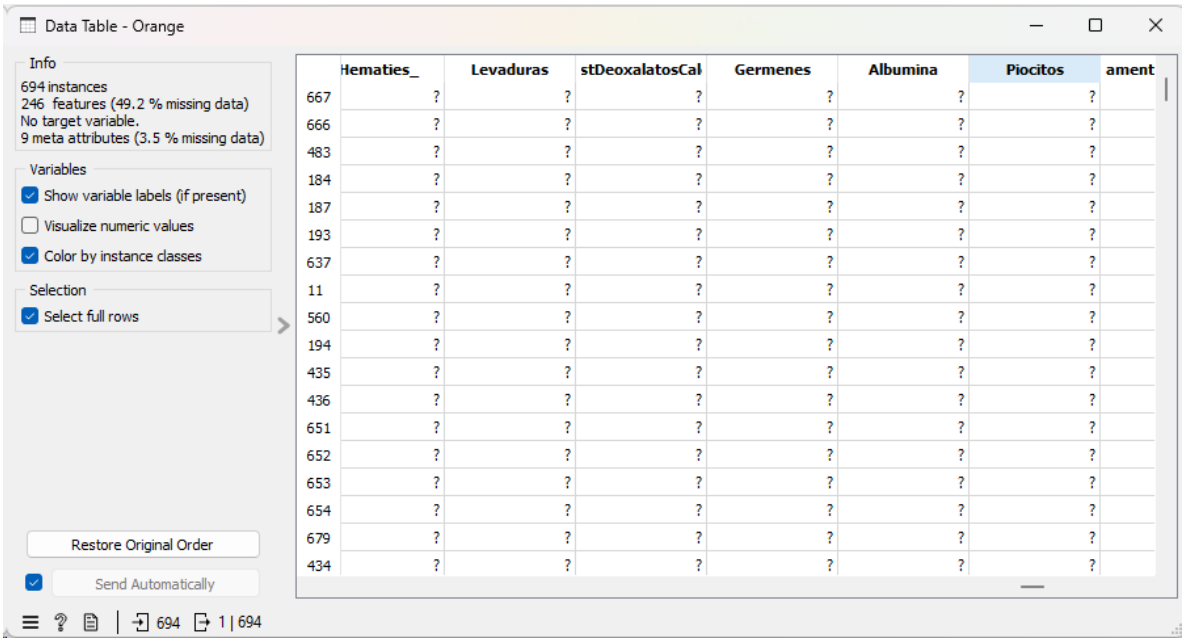


Fig. 9. Características vacías de la base de datos

Habiendo comprobado que existen muchas características que no cuentan con datos, se realizó una evaluación de todos para determinar las características iniciales que se deban seleccionara y eliminar las características que no cuenta con datos. para lograr ello se utilizó el widget select columns como se presenta en la Fig. 10.

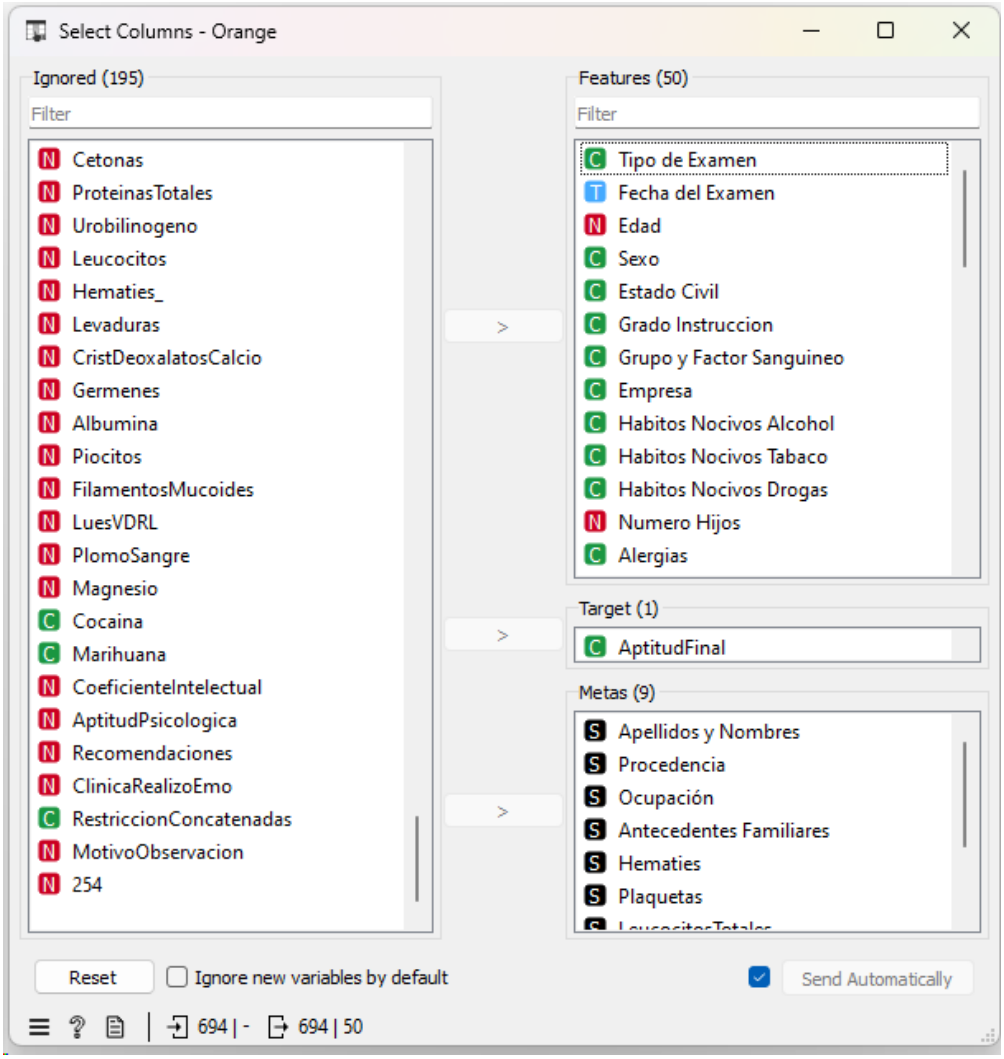


Fig. 10. Selección de las características

En la Fig. 10, se puede apreciar las características seleccionadas con el widget select columns, en total se han seleccionado 50 características (Features), se ha seleccionado una para el objetivo (Target-AptitudFinal), se han ignorazdo 195 características y 9 datos Metas.

Fig. 11 Base de datos filtrada

En la Fig. 11, se tiene la nueva base de datos con solo 50 características, el target u objetivo (con 5 tipos de valores). Se aprecia igualmente en la información, que existe 8.7% de datos perdidos o extraviados y que deberán ser tratados en la preparación de los datos previo un análisis de estos.

El modelo de minería utilizado para esta evaluación previa se muestra en la Fig. 12.

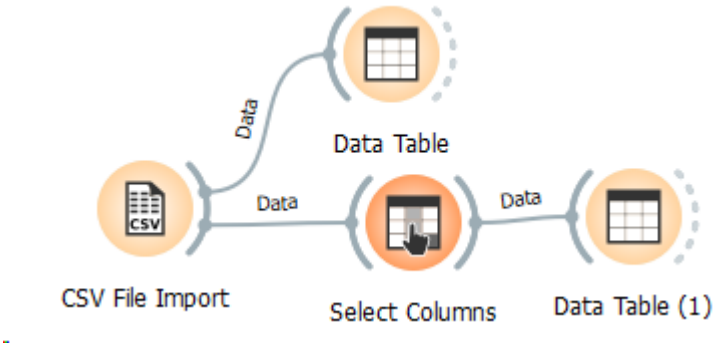


Fig. 12. Modelo de evaluación previa.

Para lograr tener una base de datos confiable para la predicción, esta debe estar libre de datos perdidos o vacíos, por lo que se debe realizar la preparación de los datos por medio de técnica de inputacion de datos.

Fase 3: Preparación de datos, Después de la recolección inicial, se seleccionan, limpian, formatean e integran los datos.

A. Preprocesado de los datos

En vista que la nueva base de datos en la que se seleccionó 50 características para evaluar, estas presentan 8.7% de datos perdidos o vacíos, se hace necesario realizar la preparación de estos datos por medio del widget preprocess como se muestra en la Fig.13

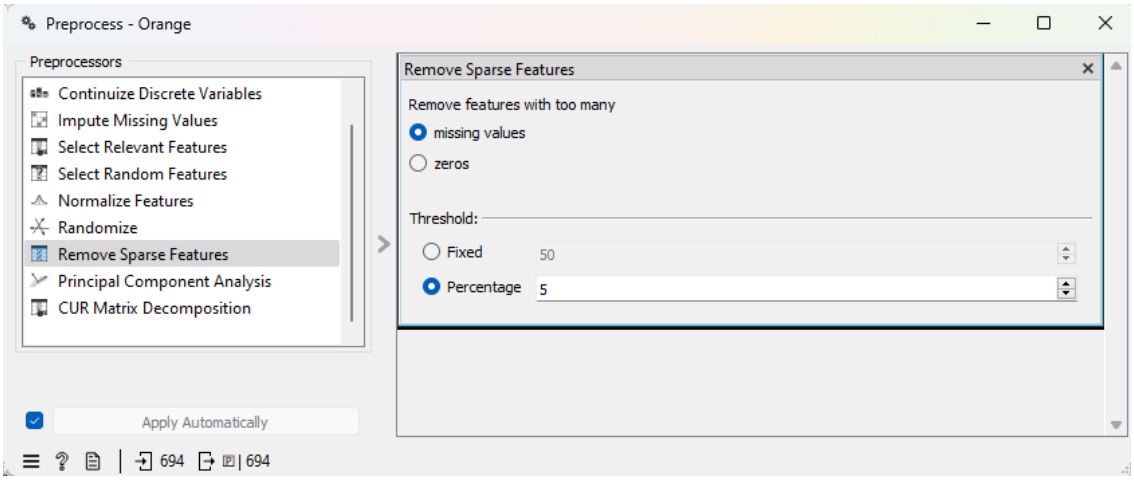


Fig. 13. Eliminación de características que tienen muchos valores perdidos

Como se aprecia en la Fig.13, con el widget de Preprocess, se selecciona el preprocesamiento inicial para eliminar aquellas características que tienen muchos datos vacíos (opción: Remove Sparce Features) en la opción missing values. Realizado este preprocesamiento inicial, la base de datos se redujo a 32 características en la que aun queda un 1.1% de valores perdidos o vacíos, como se presenta en la información de la Fig. 14.

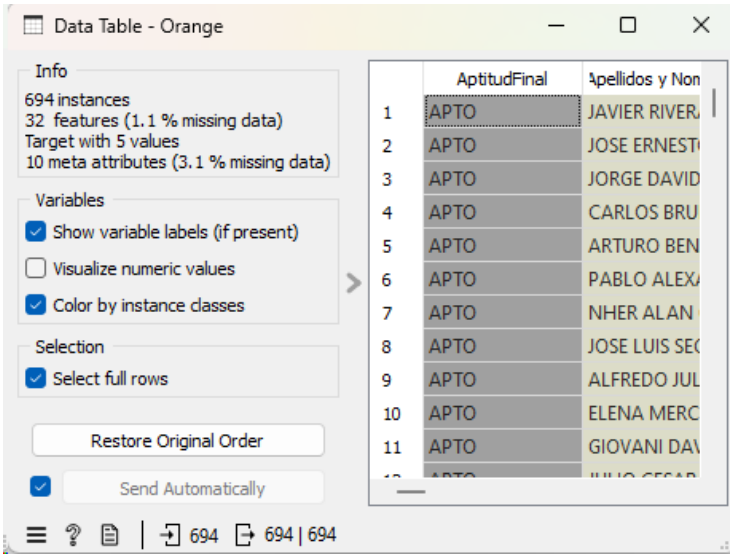


Fig. 14. Información de la base de datos con preprocesado inicial

Debido a la existencia aún de datos perdidos (1.1% missing data), se procede a realizar la inputacion de los datos para eliminar este porcentaje de datos perdidos.

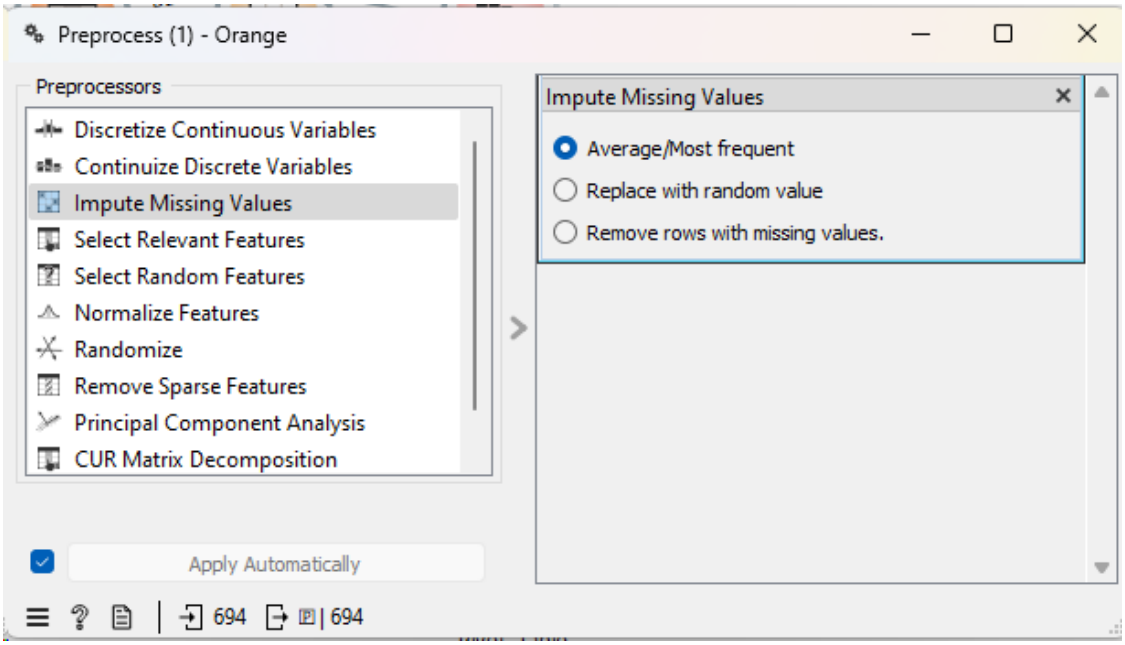


Fig. 15. Inputacion de datos

La inputacion de datos se realiza cuando se tienen datos faltantes o perdidos, las técnicas de la inputacion, se muestran en la Fig. 15 en la que se puede tomar el valor promedio o el mas frecuente que es el utilizado en el presente preprocesado de datos. También puede utilizarse el reemplazo con un valor aleatorio o simplemente eliminar las filas que tengan esos datos faltantes.

con este segundo preprocesado de datos, se tiene ahora una base de datos más sólida en la que ya no se encuentran valores perdidos en los datos, como se muestra en la Fig.16.

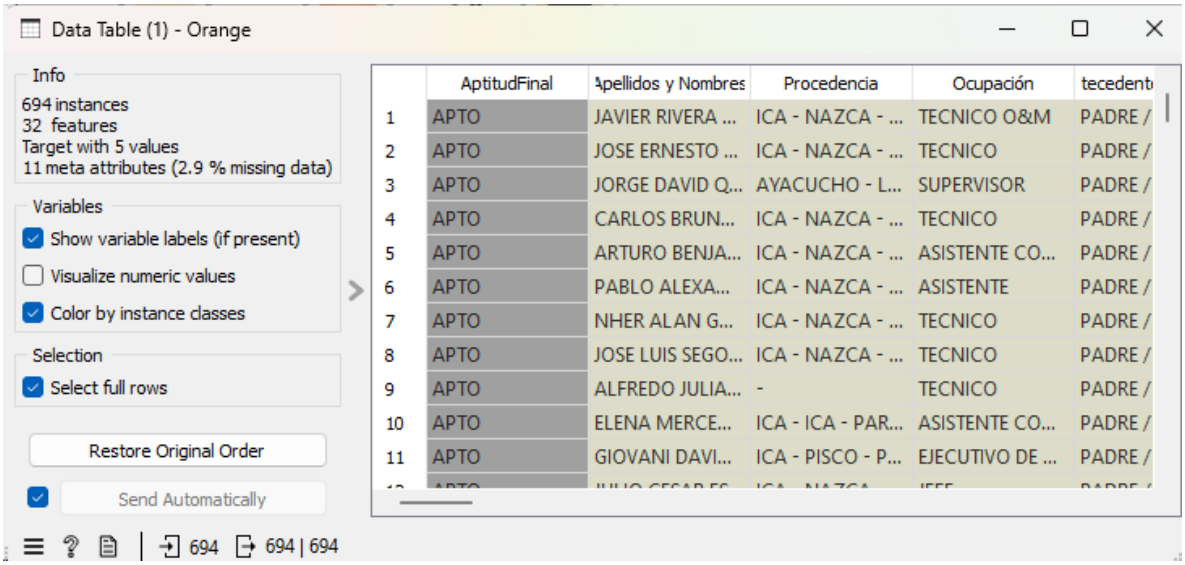


Fig. 16. Información de los datos después del 2da prepocesamiento

La información de la Fig. 16 muestra que las 32 características que han quedado, ya no tienen valores perdidos, conservando los 694 registros de datos.

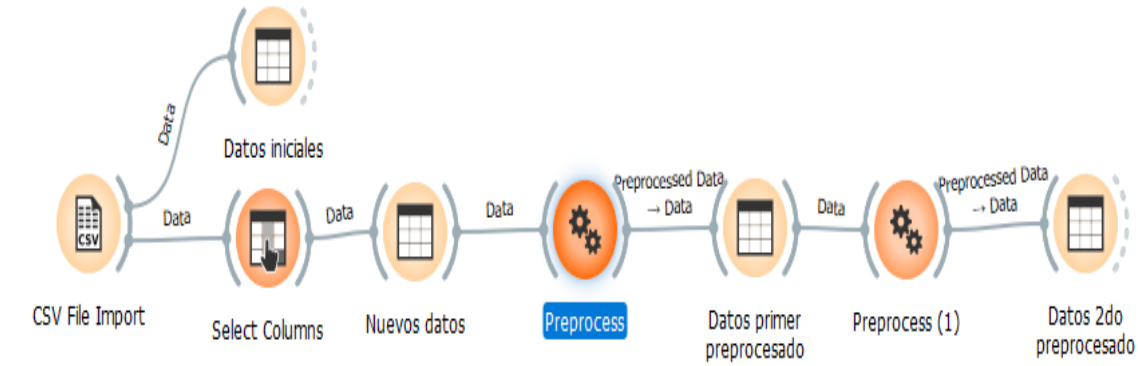


Fig. 17. Modelo orange para el preprocesado de los datos

En la Fig. 17, se muestra el procesamiento de los datos con el widget Preprocess, se tiene dos widgets para cada uno de los preprocesamiento.

Con la base de datos limpia y completa, nos quedamos con 32 características finales que se muestran en la Fig. 18. Y que se guardaron para tener un respaldo de los datos limpios y no iniciar desde cero nuevamente, por lo que a partir de estos datos se realizan la operaciones.

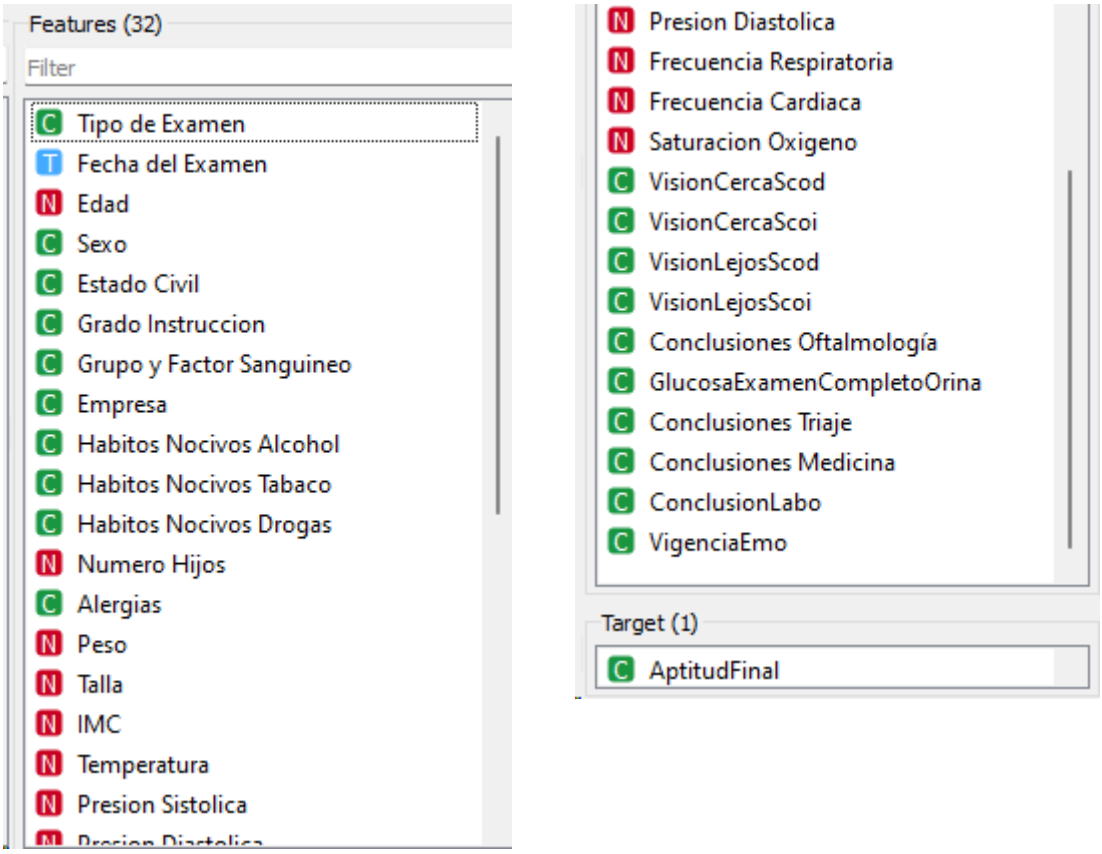
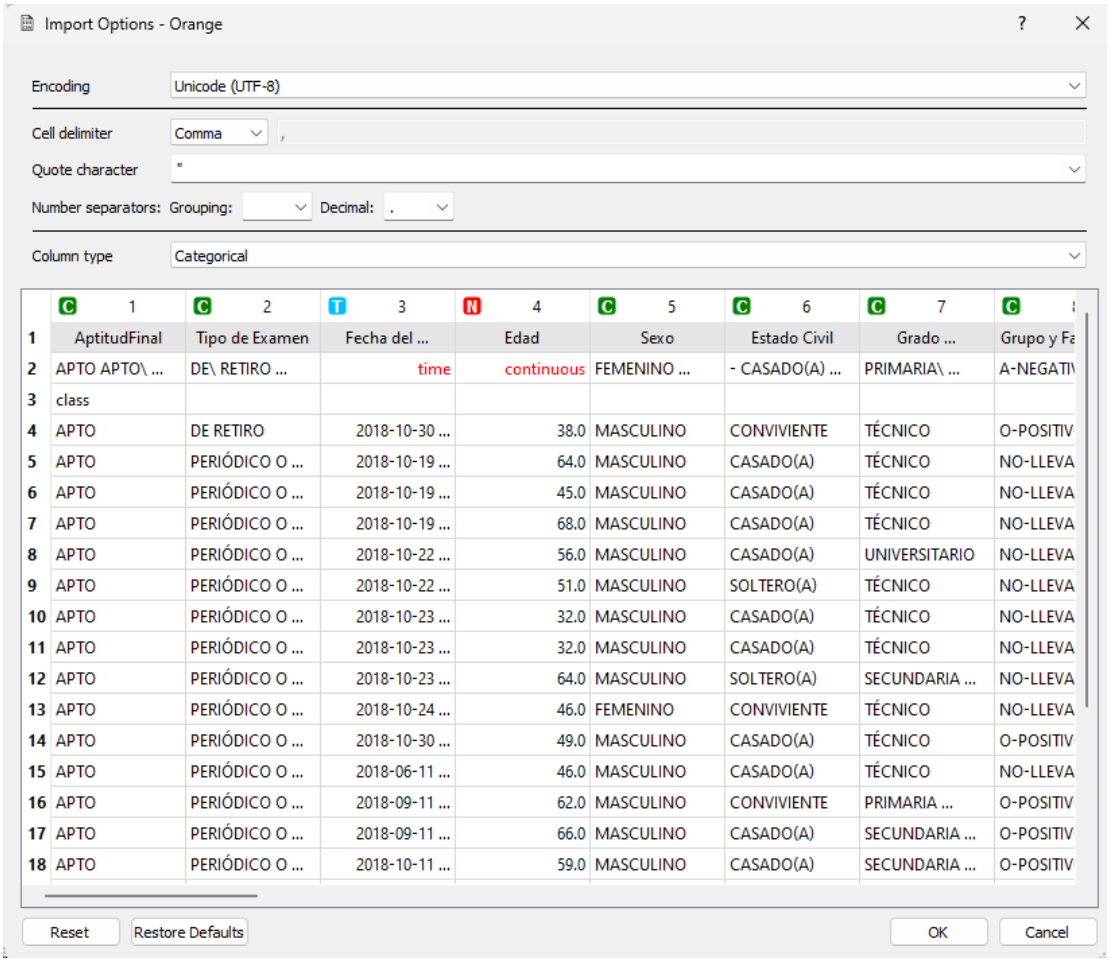


Fig. 18. Características (Features) finales

En la Fig. 18, se muestran las características finales de la base de datos limpia y sin faltantes. Además se tiene una característica destino (target) que representa la característica a evaluar con la predicción. Estos datos son guardados en un nuevo archivo Matriz de datos final.csv.

B. Análisis de los datos

Con los datos finales, desarrollamos un nuevo modelo de organge para su tratamiento, se carga el archivo de formato CSV con el widget CSV File Import, y le damos configurarmos los datos con sus tipos de datos como se muestra en la Fig. 19.



	1	2	3	4	5	6	7	
1	AptitudFinal	Tipo de Examen	Fecha del ...	Edad	Sexo	Estado Civil	Grado ...	Grupo y Fa
2	APTO APTO\ ...	DE\ RETIRO ...	time	continuous	FEMENINO ...	- CASADO(A) ...	PRIMARIA\ ...	A-NEGATI
3	class							
4	APTO	DE RETIRO	2018-10-30 ...	38.0	MASCULINO	CONVIVIENTE	TÉCNICO	O-POSITIV
5	APTO	PERIÓDICO O ...	2018-10-19 ...	64.0	MASCULINO	CASADO(A)	TÉCNICO	NO-LLEVA
6	APTO	PERIÓDICO O ...	2018-10-19 ...	45.0	MASCULINO	CASADO(A)	TÉCNICO	NO-LLEVA
7	APTO	PERIÓDICO O ...	2018-10-19 ...	68.0	MASCULINO	CASADO(A)	TÉCNICO	NO-LLEVA
8	APTO	PERIÓDICO O ...	2018-10-22 ...	56.0	MASCULINO	CASADO(A)	UNIVERSITARIO	NO-LLEVA
9	APTO	PERIÓDICO O ...	2018-10-22 ...	51.0	MASCULINO	SOLTERO(A)	TÉCNICO	NO-LLEVA
10	APTO	PERIÓDICO O ...	2018-10-23 ...	32.0	MASCULINO	CASADO(A)	TÉCNICO	NO-LLEVA
11	APTO	PERIÓDICO O ...	2018-10-23 ...	32.0	MASCULINO	CASADO(A)	TÉCNICO	NO-LLEVA
12	APTO	PERIÓDICO O ...	2018-10-23 ...	64.0	MASCULINO	SOLTERO(A)	SECUNDARIA ...	NO-LLEVA
13	APTO	PERIÓDICO O ...	2018-10-24 ...	46.0	FEMENINO	CONVIVIENTE	TÉCNICO	NO-LLEVA
14	APTO	PERIÓDICO O ...	2018-10-30 ...	49.0	MASCULINO	CASADO(A)	TÉCNICO	O-POSITIV
15	APTO	PERIÓDICO O ...	2018-06-11 ...	46.0	MASCULINO	CASADO(A)	TÉCNICO	NO-LLEVA
16	APTO	PERIÓDICO O ...	2018-09-11 ...	62.0	MASCULINO	CONVIVIENTE	PRIMARIA ...	O-POSITIV
17	APTO	PERIÓDICO O ...	2018-09-11 ...	66.0	MASCULINO	CASADO(A)	SECUNDARIA ...	O-POSITIV
18	APTO	PERIÓDICO O ...	2018-10-11 ...	59.0	MASCULINO	CASADO(A)	SECUNDARIA ...	O-POSITIV

Fig. 19. Carga de datos finales

En el widget para la carga de los datos finales, cada una de las características (o columna) se les asigna el tipo de datos al que corresponde pudiendo ser datos categoricos, numéricos y de Fecha como se muestra en la Figura. Esto se realiza con la opción Column type.

B1. Análisis preliminar de los datos

En los datos existen datos categoricos, de fecha y numericos, para el análisis preliminar de los datos numéricos se realiza el análisis descriptivo de estos datos numèricos, por medio del widget Features statistics.

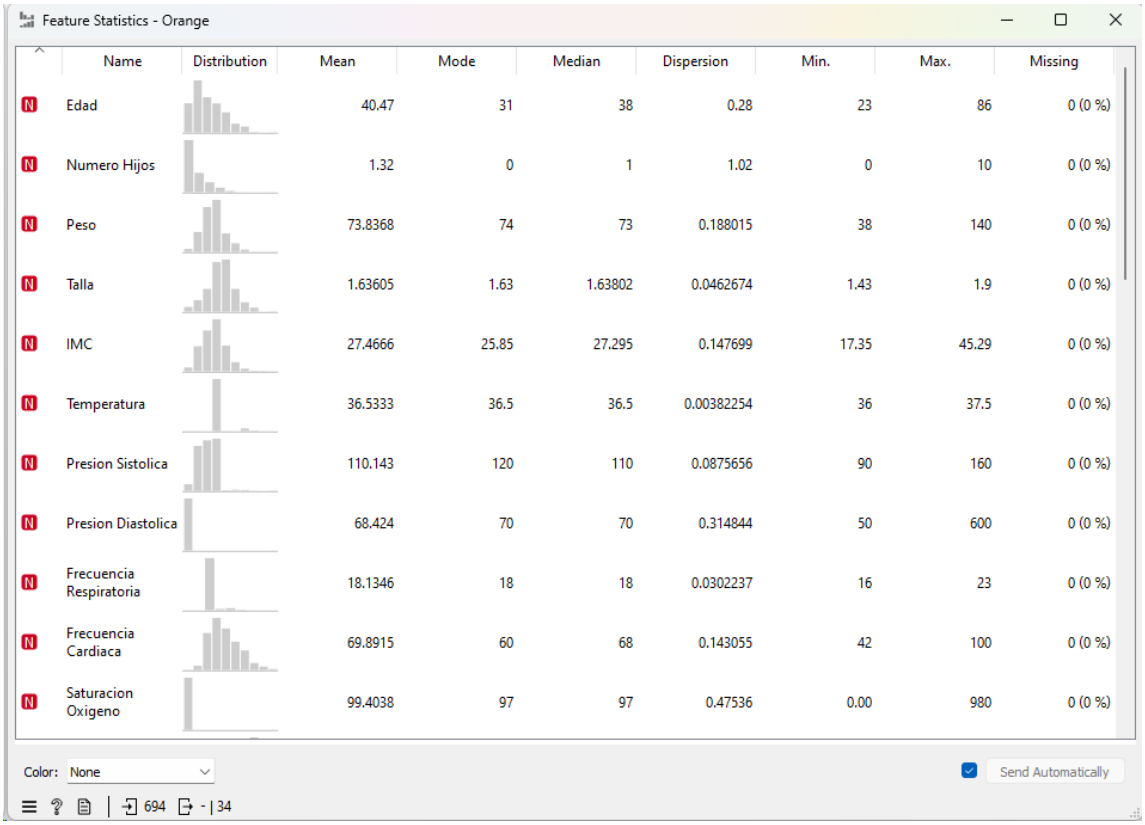


Fig. 20. Análisis descriptivo de los datos numéricos

El análisis descriptivo de los datos numéricos como se muestra en la Fig. 20, con el widget Features statistics, presenta la media, la moda, la mediana, la dispersión de los datos y los valores maximos y mínimos de cada uno de ellos.

Complementariamente con el uso del widget Pivote Table, se realiza el análisis descriptivo de los datos categoricos como se muestra en la Fig. 21.

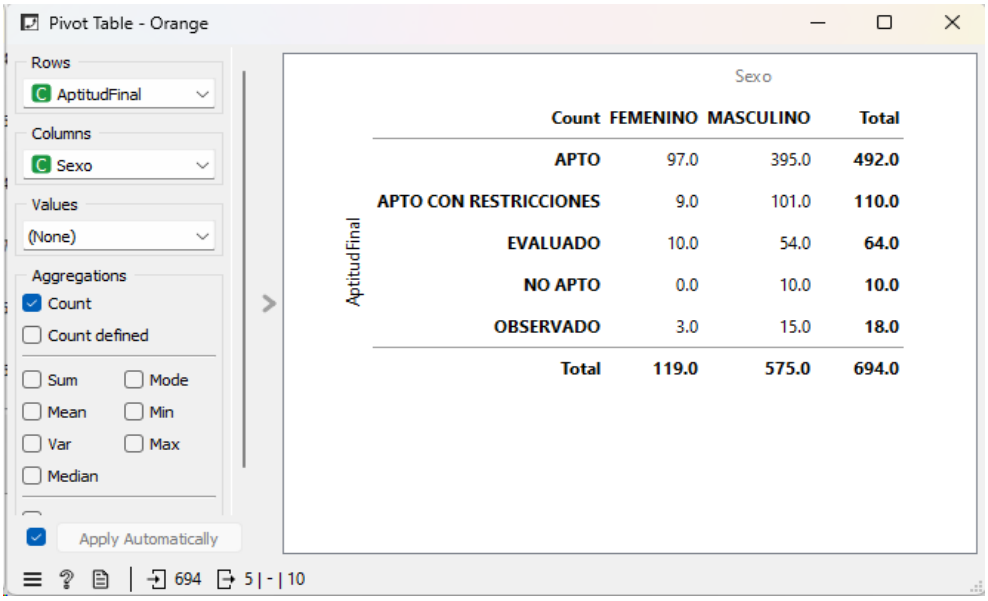


Fig. 21. Widget para el análisis de datos categoricos

En la parte izquierda del widget, se tiene los parametros de configuración. Se selecciona la característica para la fila (Rows) y la característica para la Columna (Columns). El calculo para estas características se realiza según lo que se desea de esas variables, por ser categoricas utilizaremos el conteo. En el caso de la Fig. 21, la aptitud del trabajador según el sexo muestra que de los 10 trabajadores no aptos, todos son del sexo masculino. Existe igual un numero considerable de trabajores (110) que a pesar de estar aptos presentan algunas restricciones mayormente del sexo masculino (101) y solo 9 del sexo femenino.

La tabla pivote también muestra que existen 10 trabajadores masculinos que no estan aptos, y observados 15 del sexo masculino y solo 3 del sexo femenino

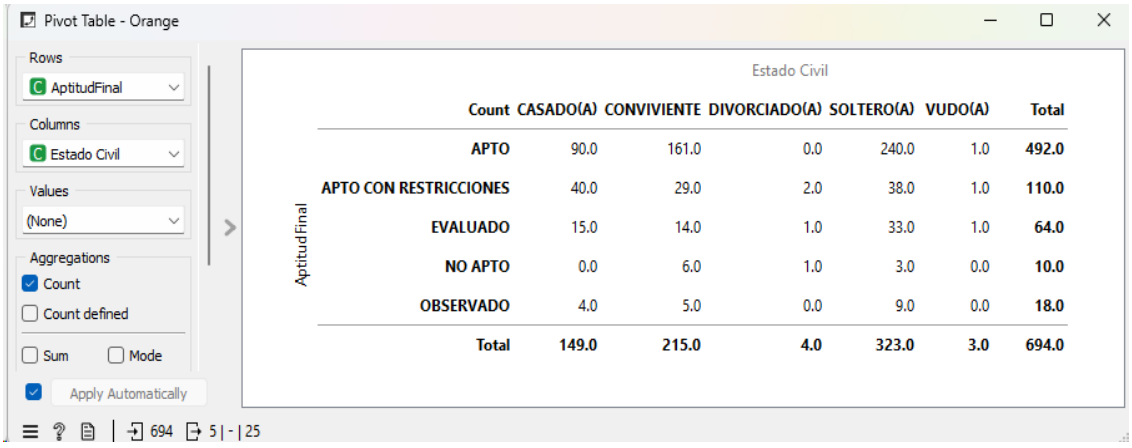
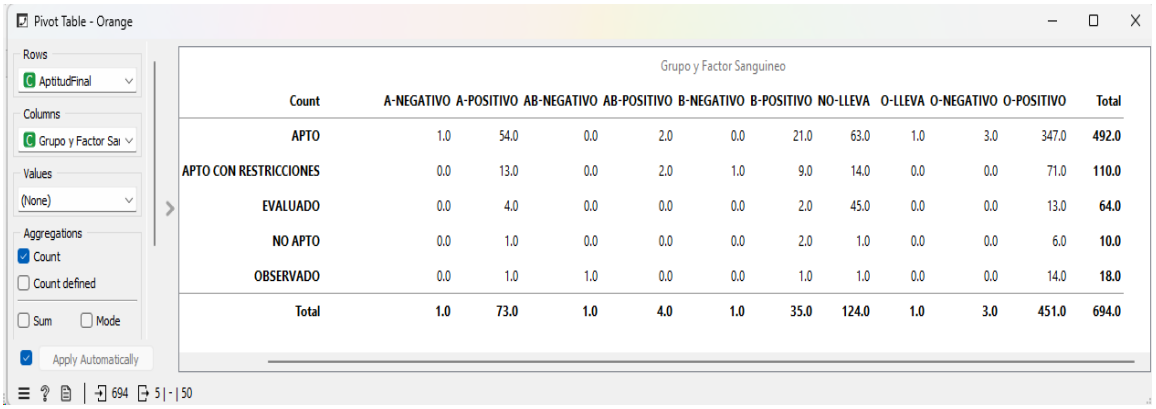


Fig. 22. Analisis de la Aptitud según el estado civil

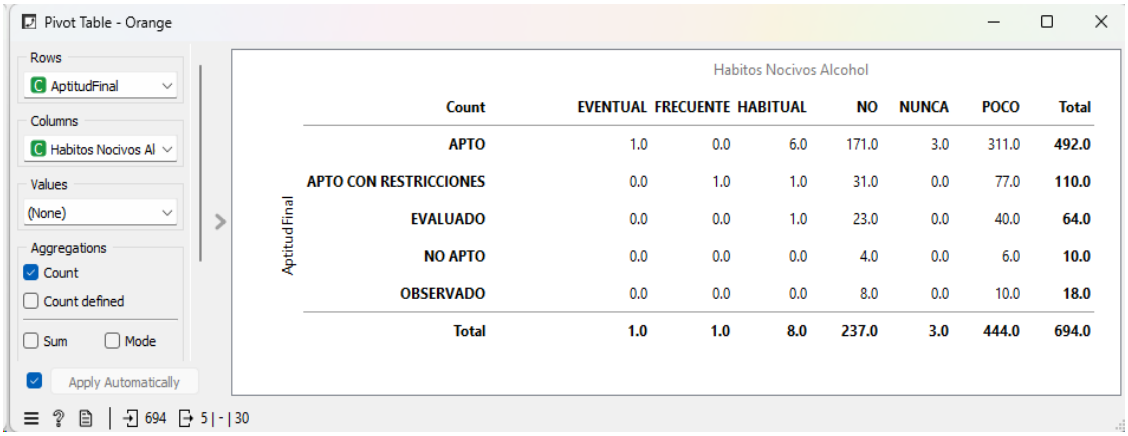
En este análisis se puede apreciar que los trabajadores aptos con restricciones, el mayor número se encuentran en trabajadores casados (40), seguido de los solteros (38). En cuanto a los no aptos estos se encuentran mayormente en los convivientes (6), y en los observados en el estado civil conviviente.



Grupo y Factor Sanguineo											
Count	A-NEGATIVO	A-POSITIVO	AB-NEGATIVO	AB-POSITIVO	B-NEGATIVO	B-POSITIVO	NO-LLEVA	O-LLEVA	O-NEGATIVO	O-POSITIVO	Total
APTO	1.0	54.0	0.0	2.0	0.0	21.0	63.0	1.0	3.0	347.0	492.0
APTO CON RESTRICCIONES	0.0	13.0	0.0	2.0	1.0	9.0	14.0	0.0	0.0	71.0	110.0
EVALUADO	0.0	4.0	0.0	0.0	0.0	2.0	45.0	0.0	0.0	13.0	64.0
NO APTO	0.0	1.0	0.0	0.0	0.0	2.0	1.0	0.0	0.0	6.0	10.0
OBSERVADO	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	14.0	18.0
Total	1.0	73.0	1.0	4.0	1.0	35.0	124.0	1.0	3.0	451.0	694.0

Fig. 23. Análisis de la aptitud según grupo sanguíneo

En la Fig. 23, se resalta en el grupo de los 110 trabajadores aptos con restricciones, el mayor número se encuentra en el grupo O-POSITIVO con 71 casos, curiosamente también el mismo grupo sanguíneo están en los No Aptos con 6, e igualmente en los observados del mismo grupo sanguíneo con 14.

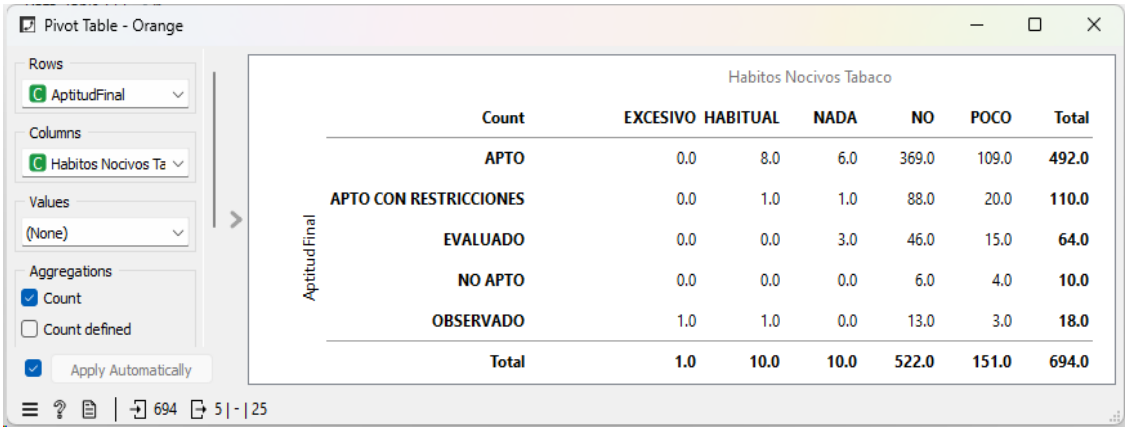


Habitos Nocivos Alcohol							
AptitudFinal / Count	EVENTUAL	FRECUENTE	HABITUAL	NO	NUNCA	POCO	Total
APTO	1.0	0.0	6.0	171.0	3.0	311.0	492.0
APTO CON RESTRICCIONES	0.0	1.0	1.0	31.0	0.0	77.0	110.0
EVALUADO	0.0	0.0	1.0	23.0	0.0	40.0	64.0
NO APTO	0.0	0.0	0.0	4.0	0.0	6.0	10.0
OBSERVADO	0.0	0.0	0.0	8.0	0.0	10.0	18.0
Total	1.0	1.0	8.0	237.0	3.0	444.0	694.0

Fig. 24. Análisis de aptitud según consumo de alcohol

Como se puede apreciar en la Fig. 24, de los 110 trabajadores aptos con restricciones, 77 han expresado tener poco consumo de alcohol. Hasta este parte del análisis es importante que esto

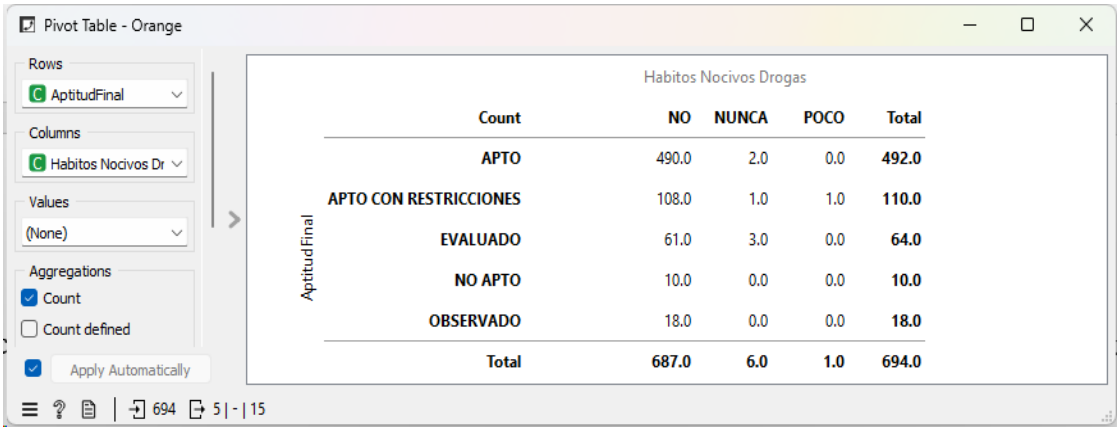
sirva para poder estructurar mejor el recojo de datos para evitar confusión en No consume o Nunca consume términos similares que pueden dar confusión.



Count	EXCESIVO	HABITUAL	NADA	NO	POCO	Total
APTO	0.0	8.0	6.0	369.0	109.0	492.0
APTO CON RESTRICCIONES	0.0	1.0	1.0	88.0	20.0	110.0
EVALUADO	0.0	0.0	3.0	46.0	15.0	64.0
NO APTO	0.0	0.0	0.0	6.0	4.0	10.0
OBSERVADO	1.0	1.0	0.0	13.0	3.0	18.0
Total	1.0	10.0	10.0	522.0	151.0	694.0

Fig. 25. Análisis de aptitud según hábitos de fumar

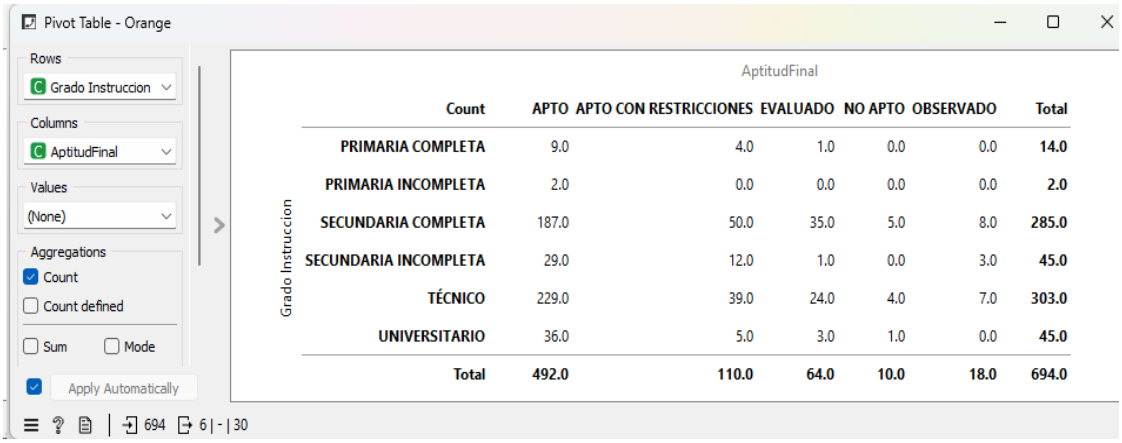
En cuanto a los hábitos de fumar, de los 110 trabajadores aptos con restricciones solo existe uno con hábitos habitual de fumar y uno en condición de observado, ninguno declarado no apto por esta característica.



Count	NO	NUNCA	POCO	Total
APTO	490.0	2.0	0.0	492.0
APTO CON RESTRICCIONES	108.0	1.0	1.0	110.0
EVALUADO	61.0	3.0	0.0	64.0
NO APTO	10.0	0.0	0.0	10.0
OBSERVADO	18.0	0.0	0.0	18.0
Total	687.0	6.0	1.0	694.0

Fig. 26. Análisis de aptitud según consumo de drogas.

En relación con los 110 trabajadores aptos con restricciones, solo uno de ello indica haber consumido drogas, pero en poca cantidad.

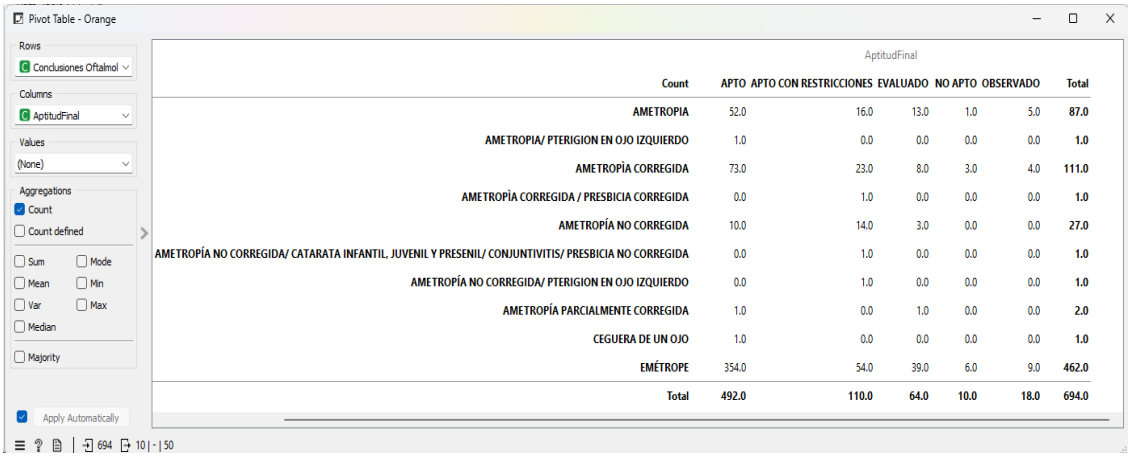


AptitudFinal

Grado Instruccion	Count	APTO	APTO CON RESTRICCIONES	EVALUADO	NO APTO	OBSERVADO	Total
	PRIMARIA COMPLETA	9.0	4.0	1.0	0.0	0.0	**14.0**
	PRIMARIA INCOMPLETA	2.0	0.0	0.0	0.0	0.0	**2.0**
	SECUNDARIA COMPLETA	187.0	50.0	35.0	5.0	8.0	**285.0**
	SECUNDARIA INCOMPLETA	29.0	12.0	1.0	0.0	3.0	**45.0**
	TÉCNICO	229.0	39.0	24.0	4.0	7.0	**303.0**
	UNIVERSITARIO	36.0	5.0	3.0	1.0	0.0	**45.0**
	Total	**492.0**	**110.0**	**64.0**	**10.0**	**18.0**	**694.0**

Fig. 27. Análisis de aptitud según su grado de instrucción

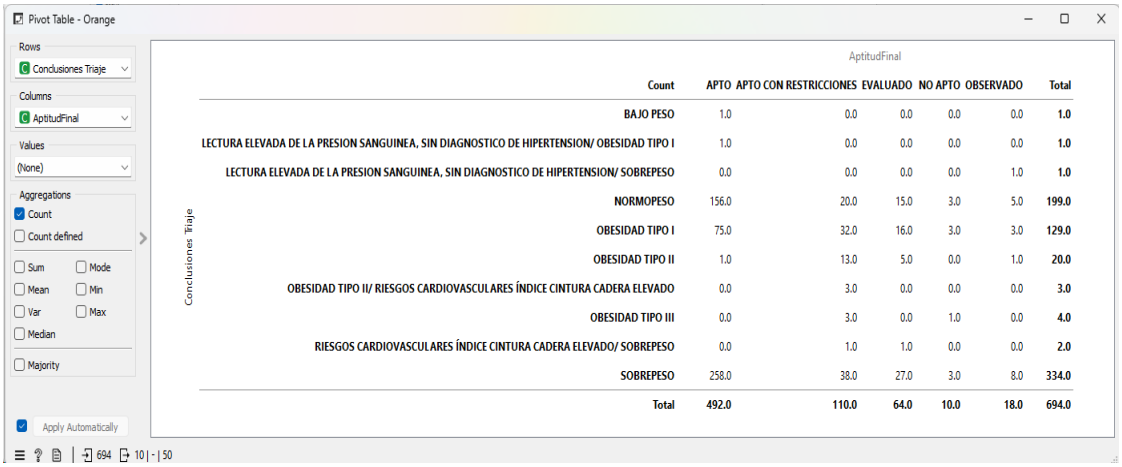
Al igual que en los análisis previos, de los 110 trabajadores aptos con restricciones existen un número importante de 50 con secundaria completa y 39 trabajadores técnicos. Y en los no aptos similarmente, pero en menor cantidad 4 de secundaria completa y 4 trabajadores técnicos. Se resalta que los trabajadores con grado universitario solo son 5 con restricciones y solo 1 no apto.



AptitudFinal

Count	APTO	APTO CON RESTRICCIONES	EVALUADO	NO APTO	OBSERVADO	Total
AMETROPIA	52.0	16.0	13.0	1.0	5.0	**87.0**
AMETROPIA/ PTERIGION EN OJO IZQUIERDO	1.0	0.0	0.0	0.0	0.0	**1.0**
AMETROPÌA CORREGIDA	73.0	23.0	8.0	3.0	4.0	**111.0**
AMETROPÌA CORREGIDA / PRESBICIA CORREGIDA	0.0	1.0	0.0	0.0	0.0	**1.0**
AMETROPÍA NO CORREGIDA	10.0	14.0	3.0	0.0	0.0	**27.0**
AMETROPÍA NO CORREGIDA/ CATARATA INFANTIL, JUVENIL Y PRESENIL/ CONJUNTIVITIS/ PRESBICIA NO CORREGIDA	0.0	1.0	0.0	0.0	0.0	**1.0**
AMETROPÍA NO CORREGIDA/ PTERIGION EN OJO IZQUIERDO	0.0	1.0	0.0	0.0	0.0	**1.0**
AMETROPÍA PARCIALMENTE CORREGIDA	1.0	0.0	1.0	0.0	0.0	**2.0**
CEGUERA DE UN OJO	1.0	0.0	0.0	0.0	0.0	**1.0**
EMÉTROPE	354.0	54.0	39.0	6.0	9.0	**462.0**
Total	**492.0**	**110.0**	**64.0**	**10.0**	**18.0**	**694.0**

Fig. 28. Análisis de aptitud según conclusiones oftalmológicas



AptitudFinal

Conclusiones Triaje — Count	APTO	APTO CON RESTRICCIONES	EVALUADO	NO APTO	OBSERVADO	Total
BAJO PESO	1.0	0.0	0.0	0.0	0.0	1.0
LECTURA ELEVADA DE LA PRESION SANGUINEA, SIN DIAGNOSTICO DE HIPERTENSION/ OBESIDAD TIPO I	1.0	0.0	0.0	0.0	0.0	1.0
LECTURA ELEVADA DE LA PRESION SANGUINEA, SIN DIAGNOSTICO DE HIPERTENSION/ SOBREPESO	0.0	0.0	0.0	0.0	1.0	1.0
NORMOPESO	156.0	20.0	15.0	3.0	5.0	199.0
OBESIDAD TIPO I	75.0	32.0	16.0	3.0	3.0	129.0
OBESIDAD TIPO II	1.0	13.0	5.0	0.0	1.0	20.0
OBESIDAD TIPO II/ RIESGOS CARDIOVASCULARES ÍNDICE CINTURA CADERA ELEVADO	0.0	3.0	0.0	0.0	0.0	3.0
OBESIDAD TIPO III	0.0	3.0	0.0	1.0	0.0	4.0
RIESGOS CARDIOVASCULARES ÍNDICE CINTURA CADERA ELEVADO/ SOBREPESO	0.0	1.0	1.0	0.0	0.0	2.0
SOBREPESO	258.0	38.0	27.0	3.0	8.0	334.0
Total	**492.0**	**110.0**	**64.0**	**10.0**	**18.0**	**694.0**

Fig. 29. Análisis de aptitud según conclusiones del triaje

C. Transformación de los datos

Habiendo realizado el análisis de los datos, se procede ahora a realizar la transformación de los datos, esto con la finalidad de tener los datos bajo un mismo formato que pueda ser utilizados por los modelos predictivos.

Se seleccionan las características relevantes de los datos, para ello se emplea el widget Select Columns en la Fig. 30 se muestran las características seleccionadas.

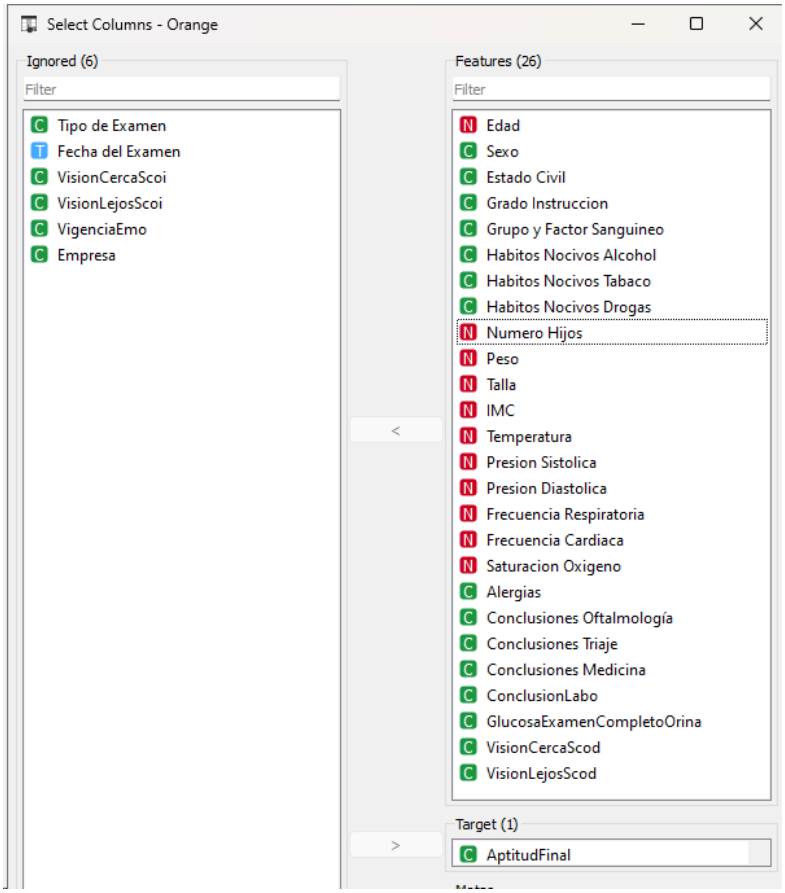


Fig. 30. Características relevantes para la predicción

Como se puede apreciar en la Fig. 30, se han seleccionado 26 características de las 32 para el proceso de predicción; las características no seleccionadas (06) no son de relevancia para la predicción. También es importante seleccionar la característica que será la que se va a predecir en nuestro caso la AptitudFinal del trabajador. Se puede apreciar igualmente que las características son de tipo numérica y de tipo categórica o texto.

Para realizar la transformación de estas características se hizo uso del widget continuize, y que se muestra en la Fig. 31.

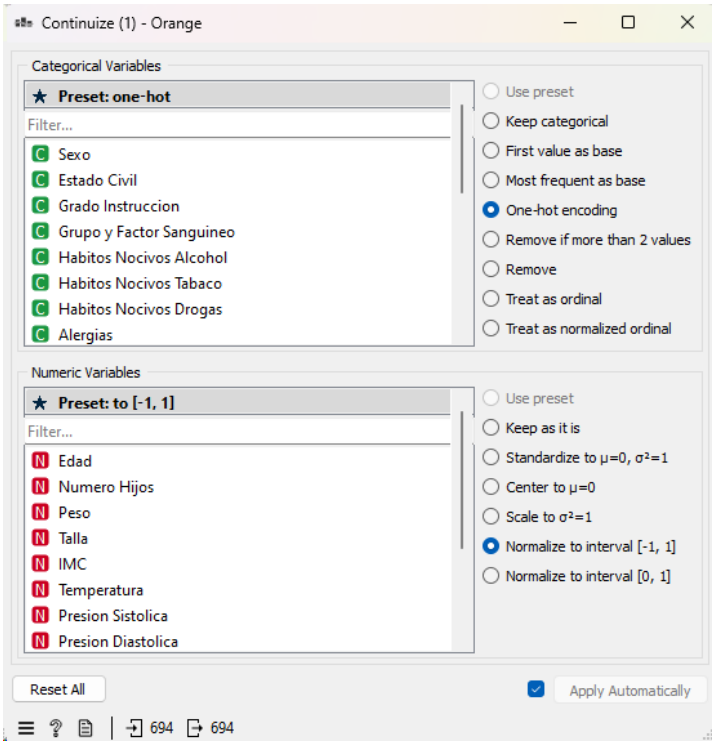


Fig. 31. Transformación de las características

El widget continuize, reconoce los tipos de características para poder realizar la transformación, en la parte superior las características de tipo categóricas y en la parte inferior las características de tipo numérica.

Para las características de tipo categórica se utiliza la técnica del one hot encoding por ser la técnica más utilizada para estas. Esta técnica permite que se transformen la categoría a numero (eg. Sexo= Masculino, Femenino en Sexo = 0,1).

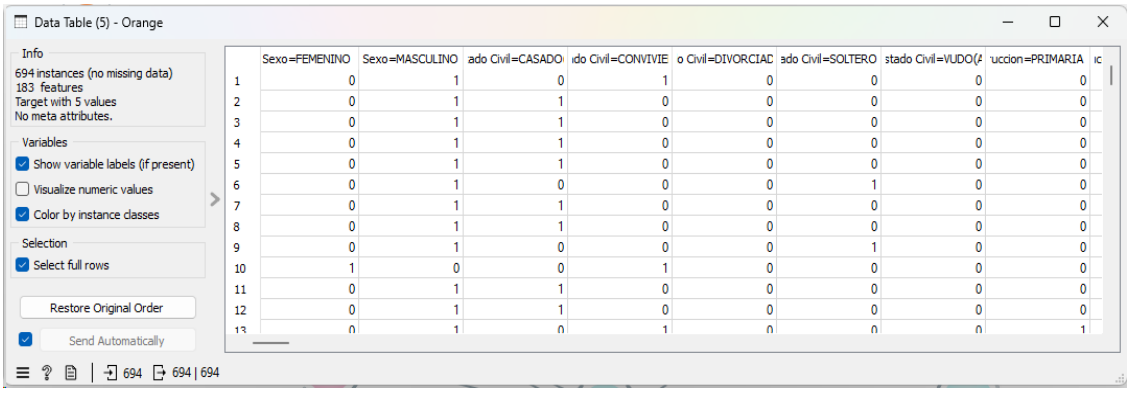
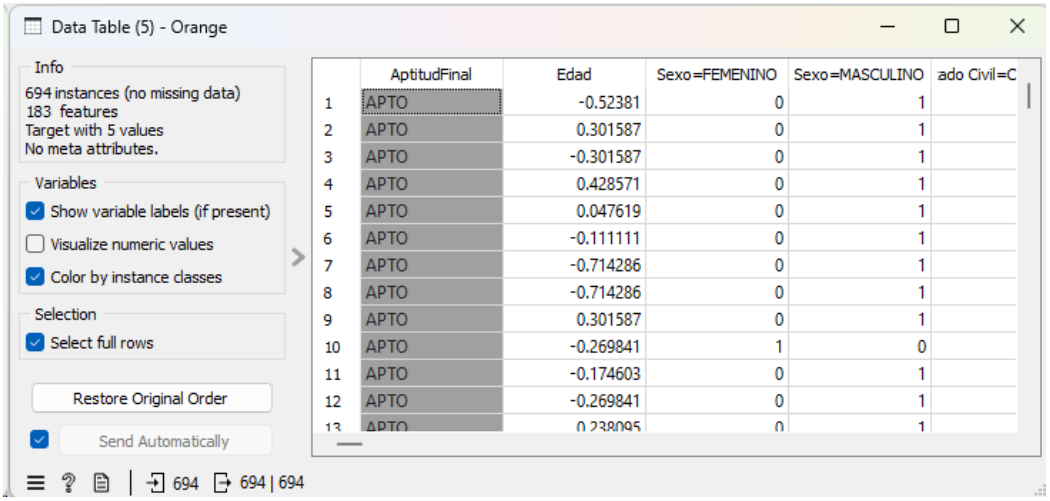


Fig. 32. Aplicación de la técnica del One hot encoding.

La Fig. 32, refleja la transformación del como se valora las opciones de la cada variable categórica.

Es importante destacar que para estas características se por cada opción o valor que contenga se genera una nueva característica por lo cual la estructura de los datos crece en función a esto, lo que se puede comprobar que de las 26 características (features) iniciales, ahora se tiene 183 características (features). Esto no sucede con las características numéricas, las que se pueden comprobar en las Figuras 33 y 34.

En el caso de las características de tipo numérica, se va a emplear la normalización de los datos para un intervalo entre [-1, 1]. Por ejemplo, la edad transformada a este intervalo de valores, como se muestra en la Fig.33.



	AptitudFinal	Edad	Sexo=FEMENINO	Sexo=MASCULINO	ado Civil=C
1	APTO	-0.52381	0	1	
2	APTO	0.301587	0	1	
3	APTO	-0.301587	0	1	
4	APTO	0.428571	0	1	
5	APTO	0.047619	0	1	
6	APTO	-0.111111	0	1	
7	APTO	-0.714286	0	1	
8	APTO	-0.714286	0	1	
9	APTO	0.301587	0	1	
10	APTO	-0.269841	1	0	
11	APTO	-0.174603	0	1	
12	APTO	-0.269841	0	1	
13	APTO	0.238095	0	1	

Fig. 33. Transformación de características numéricas

Como se puede comprobar con la característica de fecha, la edad se ha transformado a valores normalizados entre -1 y 1. En la Fig. 34, se muestra otras características transformadas.

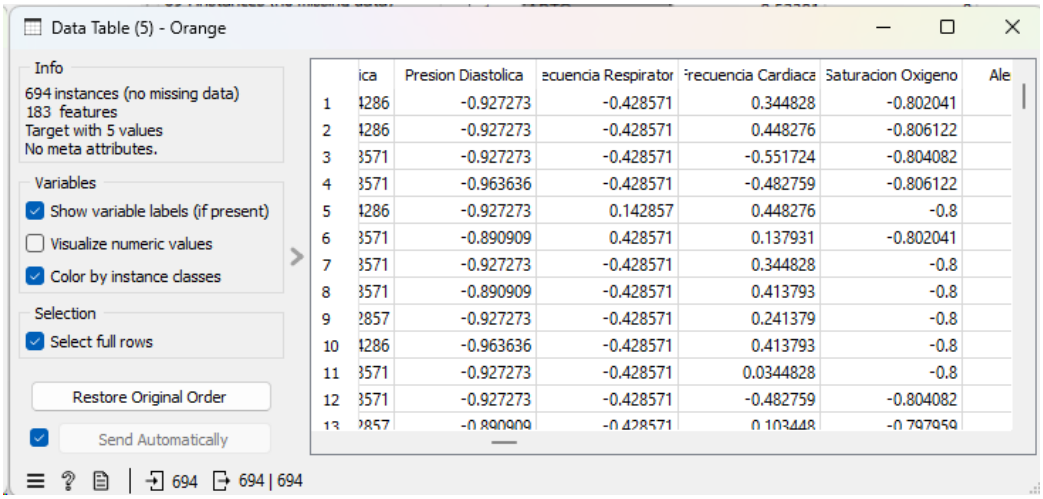


Fig. 34. Otras características transformadas

El modelo de la Fig. 35, corresponde al análisis y transformación de los datos partiendo del widget File – Select Columns – Continuize – Data Table

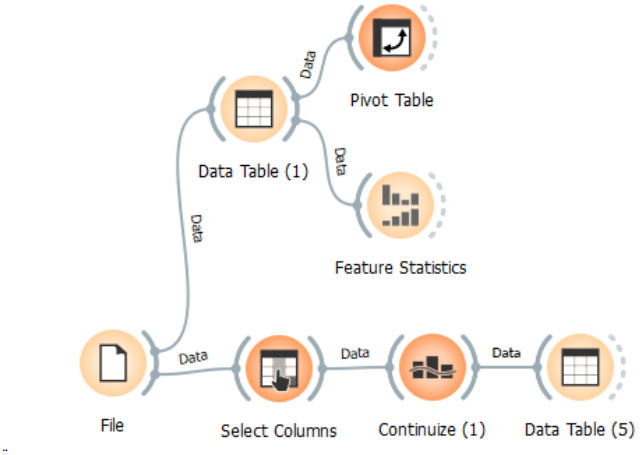


Fig. 35. Modelo para el análisis y transformación de los datos

1. File= Carga de los datos
2. Select Columns = Selecciona las características relevantes
3. Continuize = Transforma los datos tanto numéricos como categóricos para estandarizar los datos que recibirán los modelos de predicción.
4. Data Table =Almacena los datos transformados.

Fase 4: Modelado En esta fase se elige la técnica en función de: disposición de datos, conocimiento de la técnica, pertinencia con el problema y cumplir con los requisitos. Se realiza un plan de prueba para generar el modelo y evaluarlo.

Orange datamining, incorpora una serie algoritmos y modelos para realizar la predicción de variables. Por lo que se realizara inicialmente la prueba de los modelos para encontrar cual es el modelo más preciso en la predicción con la que finalmente se obtendrán los resultados. Estos modelos se presentan en la Fig. 36.

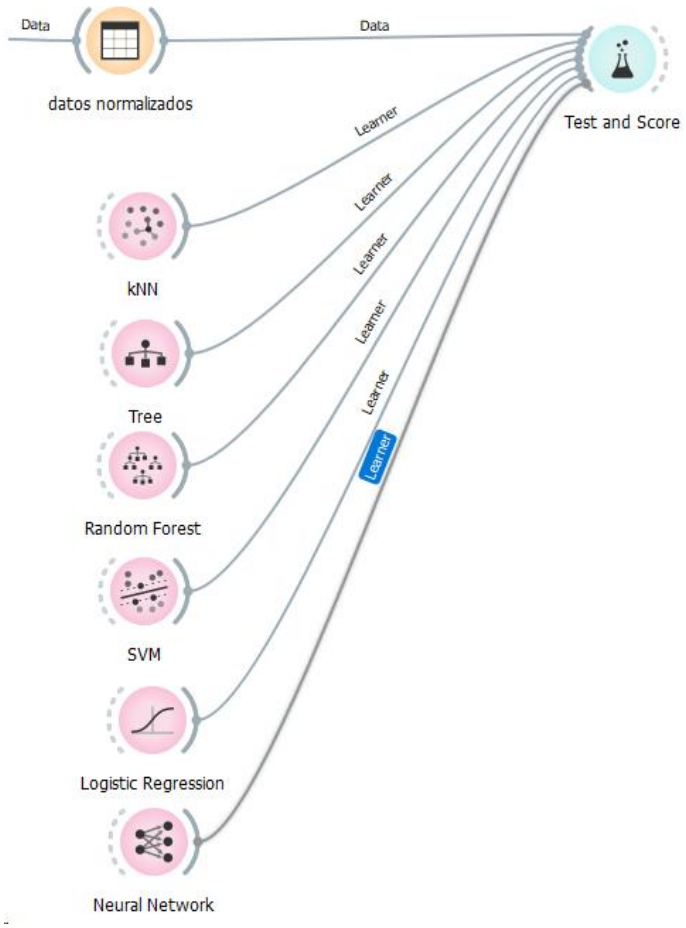


Fig. 36. Modelos de predicción para la característica AptitudFinal

Como se muestra en la Fig. 36, el modelo se esta realizando una prueba de 6 modelos para la clasificación con los widgets KNN, Tree, Random Forest, SVM, Logist Regression y Neural Network. Para hacer una evaluación de cada uno de los modelos se utiliza el widget Test and Score, al que se le ingresa los datos normalizados y los algoritmos de los modelos.

El resultado del Test and Score, aplica una serie de métricas para ver cual de los modelos es el mas eficiente para la data ingresadas, los resultados se muestran en la Fig. 37.

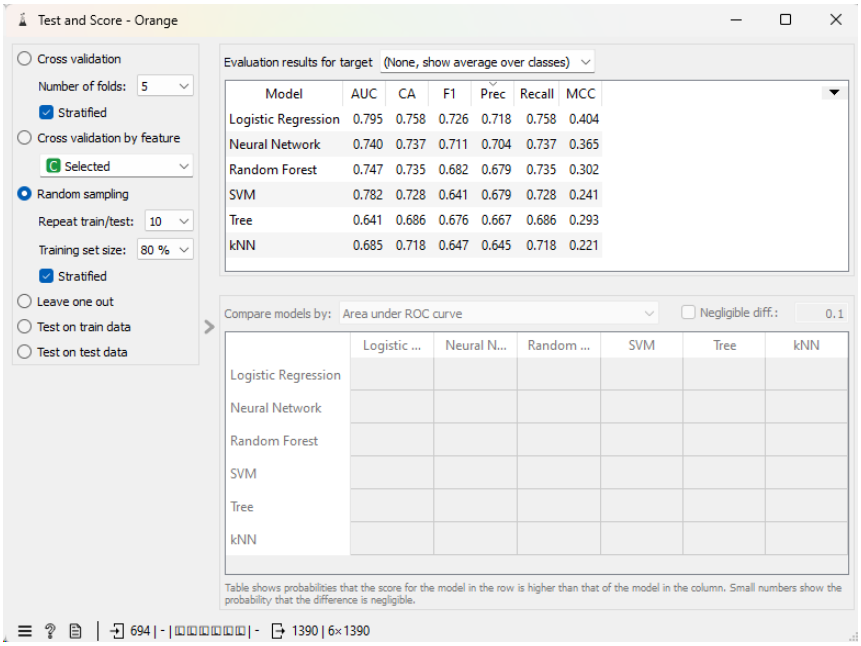


Fig. 37. Resultado de las métricas

Para realizar un mejor análisis de las métricas, se ha ordenado los resultados de mayor a menor, lo cual se realiza con un simple clic en el título de la métrica (en este caso la precisión).

Según la precisión de la predicción se puede comprobar que el modelo de la regresión logística presenta la mejor precisión con un 75,80%, seguido de la red neuronal con 73,70%, utilizando el 80% de los datos para el entrenamiento y 20% para el test. Se podrá realizar diversas pruebas con los valores para el entrenamiento a voluntad del investigador y determinar finalmente cual será el mejor modelo de predicción para los datos.

Realizada diversas pruebas con diversas muestras que se selecciona se obtuvo los siguientes resultados para datos de entrenamiento del 75%., en la que se obtuvo una precisión del 75,5% y área bajo la curva AUC 0.795 el más cercano a 1a unidad. Los resultados se muestran en la Fig. 38.

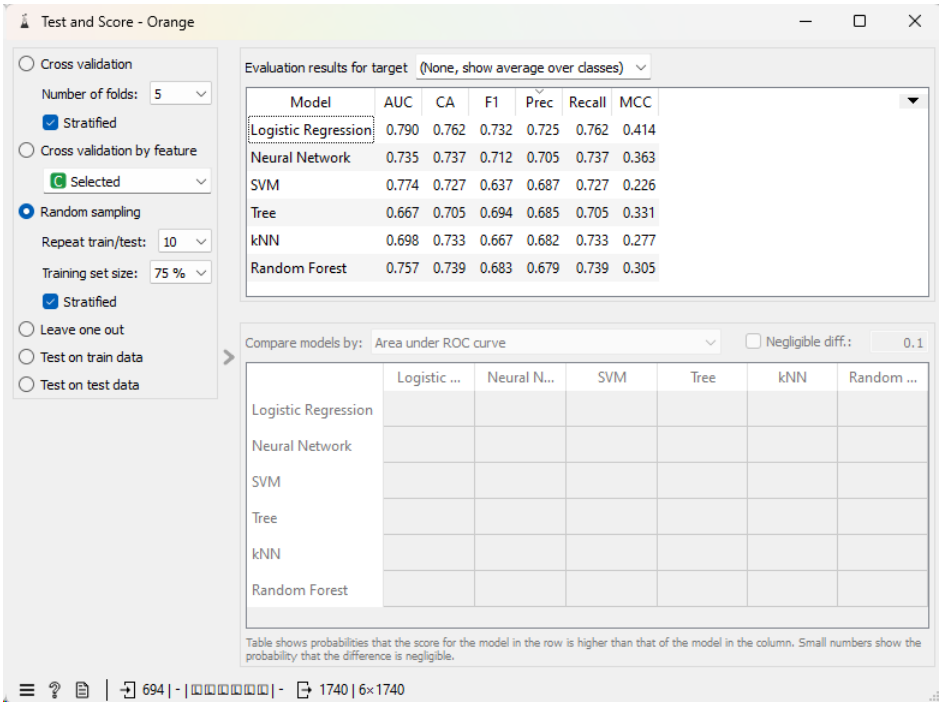


Fig. 38. Resultados con 75% de datos de entrenamiento

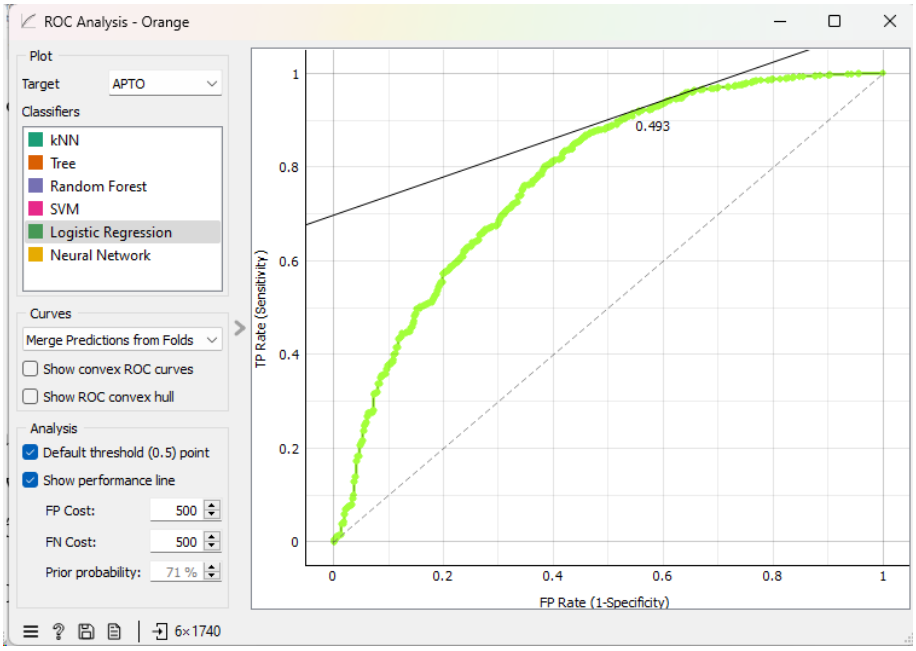


Fig. 39. Gráfica de la curva AUC

La gráfica de la curva AUC, es un indicativo de la predicción, cuanto más cercana este a la unidad es mejor la predicción. La grafica muestra que hay una cercanía a la unidad para la categoría APTO de la AptitudFinal.

En la grafica puede realizarse diversas variaciones solo cambiando el modelo de predicción, como también la categoría del target.

Seleccionado el modelo con mayor precisión se realiza el procesamiento complementario solo con el modelo de regresión logística, se puede ajustar los parámetros de este modelo. El algoritmo para la regresión logística puede ser configurando el tipo de regularización a L1 o a L2, estos cambios modifican el resultado de la precisión L2(Ridge)=precisión 79.5%, y L1(Lasso)=78.5%.

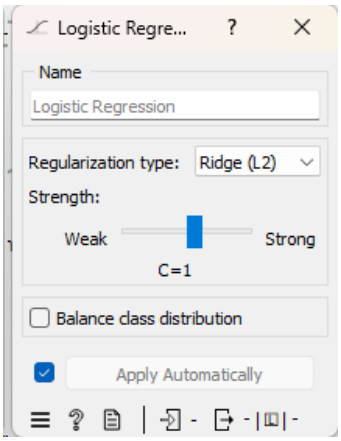


Fig. 40. Configuración del algoritmo de regresión logística

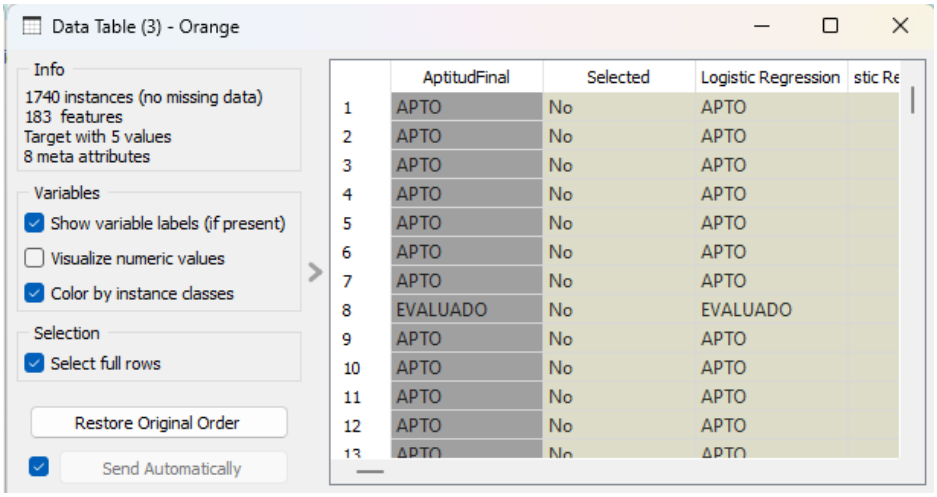


Fig. 41. Resultados de la predicción.

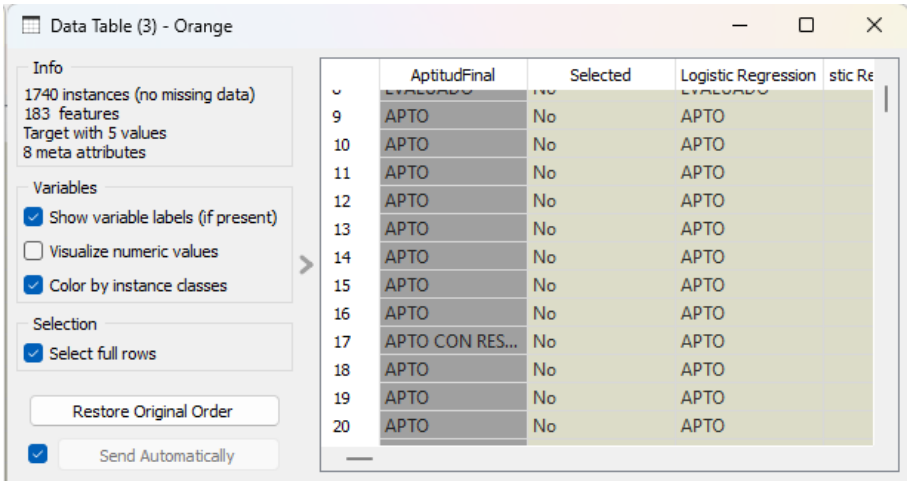


Fig. 42. Resultados de la predicción

En las Figuras 41 y 42, se aprecia los resultados de la predicción, con el nivel de precisión de cerca del 80%.

Fase 5: Evaluación, Se evalúan los resultados (Capitulo III), se revisa el proceso para identificar deficiencias. Si el modelo generado es válido en función de los criterios establecidos en la fase anterior, se procede a la implementación del Modelo.

Siendo válido el modelo por los resultados obtenidos de más de 70% de precisión, se identifica que la cantidad de datos no es lo suficientemente grande para obtener predicciones más precisas.

Fase 6: Implementación, En esta fase se propone contar con estrategias para monitorear y mantener el modelo. En el desarrollo de la tesis esta etapa, se considera una limitación porque no es posible realizar la implementación en el policlínico, hasta que no se consiga una mayor precisión del modelo, por lo que se requiere ampliar la cobertura de los datos a todos los datos históricos.

III. RESULTADOS

3.1 Descripción de los trabajadores evaluados

A. Descripción de variables categóricas

TABLA I
TRABAJADORES EVALUADOS SEGÚN SEXO

SEXO	Frecuencia	%
FEMENINO	119	17,15
MASCULINO	575	82,85
	694	100,00

De los 694 trabajadores evaluados en el policlínico San pablo como se ve en la tabla mayormente son del sexo masculino con cerca del 83% y solo el 17% del sexo femenino.

TABLA II
TRABAJADORES EVALUADOS SEGÚN ESTADO CIVIL

ESTADO CIVIL	Frecuencia	%
CASADO	149	21,47
CONVIVIENTE	215	30,98
DIVORCIADO	4	0,58
SOLTERO	323	46,54
VIUDO	3	0,43
	694	100,00

En el caso de los 694 trabajadores evaluados en relación con su estado civil, se desprende de la tabla una mayor proporción de trabajadores solteros con 46,54%, seguido de los convivientes con un 30,98%. la cantidad de trabajadores casados solo representan un 21,47%.

TABLA III
TRABAJADORES EVALUADOS SEGÚN SEXO Y ESTADO CIVIL

ESTADO CIVIL	FEMENINO	%	MASCULINO	%	TOTAL	%
CASADO	16	2,31	133	19,16	149	21,47
CONVIVIENTE	22	3,17	193	27,81	215	30,98
DIVORCIADO	1	0,14	3	0,432	4	0,58
SOLTERO	79	11,38	244	35,16	323	46,54
VIUDO	1	0,14	2	0,288	3	0,43
	119	17,15	575	82,85	694	100,00

De los 694 trabajadores evaluados, el mayor porcentaje corresponde a los solteros masculinos con un 35.16% seguido de los convivientes masculinos con un 27,81%. en el caso de las mujeres mayormente son solteras con un 11,38%.

TABLA IV
TRABAJADORES EVALUADOS SEGÚN SEXO DE LAS PRINCIPALES EMPRESAS

EMPRESA	FEMENINO	%	MASCULINO	%	TOTAL	%
APC CORPORACION S.A.	71	19,94	99	27,81	170	47,75
CANTALLOC S.R.L.	2	0,56	36	10,11	38	10,67
EMAPAVIGS S.A.	7	1,97	27	7,58	34	9,55
AUTORIDAD NACIONAL DEL AGUA	5	1,40	18	5,06	23	6,46
V Y T CONTRATISTAS SAC	0	0,00	22	6,18	22	6,18
TERRAMOVE SAC	10	2,81	59	16,57	69	19,38
	95	26,69	261	73,31	356	100,00

Para el caso de los trabajadores según la empresa donde labora, se seleccionó las empresas que tienen la mayor cantidad de trabajadores evaluados que son presentadas en la Tabla IV, y que en total son 356 poco más del 50% de todos los trabajadores. La mayor cantidad de trabajadores evaluados corresponde a la empresa APB CORPORACION SA con 27,81% del sexo masculino y 19,94 del sexo femenino. La empresa TERRAMOVE SAC, con la segunda

mayor proporción de trabajadores evaluados con 16,57 del sexo masculino y solo 2,81 del sexo femenino.

B. Descripción de variables numéricas

TABLA V
MEDIDAS DE TENDENCIA CENTRAL DE CARACTERISTICAS NUMERICAS

Categoría	Media	Mediana	Moda
Edad	40,47	38	31
Número de hijos	1,32	1	0
Peso	73,84	73	74
Talla	1,64	1,64	1,63
IMC	27,47	27,3	25,85
Temperatura	36,53	36,5	36,5
Presión Sistólica	110,14	110	120
Presión Diastólica	68,42	70	70
Frecuencia respiratoria	18,13	18	18
Frecuencia cardiaca	69,89	68	60
Saturación de Oxigeno	99,40	97	97

En la Tabla V, en las que se describen las características numéricas. La tabla muestra que los trabajadores evaluados tienen una edad media de 40,47 años, con un 01 hijo y un peso de 73,84 kilos y una talla media de 1,64 mts. Una variable importante en esta descripción es el IMC (Índice de Masa Corporal) con un índice de 27.47, además presentan una temperatura media de 36.53 ºC.

La presión media de los trabajadores evaluados está en el rango de 110 – 68, con una saturación de oxígeno de 99%.

3.2 Estado de la Aptitud de los trabajadores

TABLA VI
ESTADO DE LA APTITUD DE LOS TRABAJADORES

Aptitud Final	Frecuencia	%
APTO	492	70,89
APTO CON RESTRICCIONES	110	15,85
EVALUADO	64	9,22
NO APTO	10	1,44
OBSERVADO	18	2,59
	694	100,00

Con relación a el estado de los trabajadores de las empresas que ha sido evaluados para determinar la aptitud para el trabajo, la tabla VI, muestra los resultados en la que más del 70% se encuentran aptos, un aproximado de 16% están aptos, pero con restricciones. Solo 1,44% de estos trabajadores no están aptos para el trabajo.

TABLA VII
ESTADO DE LAS APTITUDES DE LOS TRABAJADORES DE LAS EMPRESAS MAS REPRESENTATIVAS

EMPRESA	APTO	%	APTO CON RESTRICCIONES	%	EVALUADO	%	NO APTO	%	OBSERVADO	%	TOTAL	%
APC CORPORACION S.A.	144	40,45	15	4,21	0	0,00	5	1,40	6	1,69	170	47,75
CANTALLOC S.R.L.	21	5,90	17	4,78	0	0,00	0	0,00	0	0,00	38	10,67
EMAPAVIGS S.A.	21	5,90	11	3,09	0	0,00	0	0,00	2	0,56	34	9,55
AUTORIDAD NACIONAL DEL AGUA	20	5,62	3	0,84	0	0,00	0	0,00	0	0,00	23	6,46
V Y T CONTRATISTAS SAC	19	5,34	3	0,84	0	0,00	0	0,00	0	0,00	22	6,18
TERRAMOVE SAC	12	3,37	3	0,84	51	14,33	3	0,84	0	0,00	69	19,38
	237	66,57	52	14,61	51	14,33	8	2,25	8	2,25	356	100,00

En la Tabla se presentan el estado de los trabajadores de las empresas que mayor cantidad de trabajadores han sido evaluados. La empresa APC CORPORACION SA cuenta con el mayor porcentaje de aptos con 40,45% de los 356 trabajadores de estas empresas.

3.3 Pronóstico de la Aptitud de los trabajadores

Con los modelos y los algoritmos utilizados para evaluar los que tienen una mayor eficiencia en la predicción de la Aptitud de los trabajadores, los mismos que se presenta en la Fig. 37. Y que para efectos de los resultados se toma las métricas de los dos mejores modelos como es la regresión logística y las redes neuronales.

TABLA VII
RESULTADOS DE LAS METRICAS

Métrica	Modelo	
	Regresión Logística	Redes Neuronales
AUC (área bajo la curva)	0,790	0,735
CA (Precisión de la clasificación)	0,762	0,737
F1	0,732	0,712
Prec (Precisión)	0,725	0,705
Recall	0,762	0,737
MCC	0,414	0,363

Los resultados de las métricas para la cantidad de datos procesados, indica una precisión de la clasificación de la Aptitud de los trabajadores del 76,2% y 73,7% para los dos mejores modelos. Se considera este valor muy valioso para una muestra pequeña como es el valor 694.

Descripción de las métricas

AUC : El área bajo ROC es el área bajo la curva operativa del receptor.

CA : La precisión de la clasificación es la proporción de ejemplos clasificados correctamente.

F1 : Es una media armónica ponderada de precisión y recuperación.

Prec : La precisión es la proporción de verdaderos positivos entre los casos clasificados como positivos.

Recall : La recuperación es la proporción de verdaderos positivos entre todas las instancias positivas de los datos.

MCC : El coeficiente de correlación de Matthews tiene en cuenta los positivos y negativos verdaderos y falsos, y suele considerarse una medida equilibrada que puede utilizarse incluso si las clases son de tamaños muy diferentes.

IV. DISCUSIÓN

La inteligencia artificial está creciendo a pasos muy grandes presentando soluciones algorítmicas en el campo del diagnóstico y tratamiento de paciente [5], con un enfoque similar a nuestro estudio en el pronóstico de la Aptitud de trabajadores. En la mayoría de proyectos de minería de datos en la que se emplea un 70% de los datos para el entrenamiento y un 30% para la prueba, con los que se obtuvo diversas precisiones de los datos desde 74,73% hasta precisiones hasta 93,70% con el modelo de Arboles de decisión[2], resultados similares en nuestro proyecto con una precisión del 76,20% para una configuración de 75% de datos para el entrenamiento y 25% para la prueba con el algoritmo de regresión logística y Redes neuronales. Diversos estudios apoyan el uso de la metodología CRISP-DM [21], [2] [22], empleada en nuestro trabajo de tesis.

V. CONCLUSIONES

Nuestro estudio de tesis se centró en el Policlínico San Pablo de la provincia de Nazca de la región Ica, debido al interés de poder evaluar la posibilidad de realizar el pronóstico de la Aptitud con los exámenes médico ocupacional para el trabajo de los trabajadores, además de hacer una descripción de estos trabajadores que acuden al policlínico, lo que el estudio evidencia es la capacidad de la Inteligencia Artificial, para realizar el análisis y predicción.

1. Se ha podido realizar mediante el software Orange dataminig la descripción de las características o pruebas más relevantes de los análisis realizados a los trabajadores de las empresas que acuden al Policlínico San Pablo de la ciudad de Nazca de forma simple y sencilla, sin necesidad de recurrid a la programación del lenguaje Python, precisando que el lenguaje detrás de los widgets del software esta el lenguaje Python. Se ha realizado la descripción de las características categóricas y características numéricas.
2. Similarmente se ha determinado la aptitud de los trabajadores de las empresas considerando las empresas más relevantes, que se corresponde con las empresas de mayor cantidad de trabajadores evaluados.
3. Los clasificadores utilizados con los que mejores resultados arrojan en el pronóstico de la aptitud de los trabajadores. A pesar de la cantidad de datos recogida del policlínico el nivel de precisión se considera bueno. sin embargo, no lo suficiente por ser de mucha delicadeza un diagnóstico a nivel médico.
4. Concluimos igualmente que no se requiere ser un experto en lenguaje de programación, para realizar análisis de datos o predicciones, debido a que la herramienta facilita la labor del análisis y predicción de los datos.

VI. RECOMENDACIONES

Con los buenos resultados obtenidos tanto en el análisis descriptivo de los datos, como en la predicción o pronóstico de la Aptitud de los trabajadores de las empresas que acuden al policlínico San Pablo, planteamos las siguientes recomendaciones:

1. Se recomienda trabajos futuros sobre el uso de la inteligencia artificial con mayor cantidad de datos históricos sobre la Aptitud de los trabajadores, además de otros tipos de diagnósticos médicos que se realicen en el policlínico San Pablo. Considerar en estas nuevas investigaciones emplear para las pruebas todos los datos históricos que tuviera el policlínico, ya que es importante destacar que cuanto mayor es la cantidad de datos que se proporcionen a los algoritmos la precisión de la predicción o pronóstico será mayor.
2. Se recomienda que el policlínico cuente con personal capacitado en el campo de la inteligencia artificial, ya que puede ser el soporte de los diagnósticos y el tratamiento de muchas enfermedades. Ideal se considera que se capacite al personal encargado de los sistemas y bases de datos para aprovechar la data registrada en sus sistemas.

VII. REFERENCIAS BIBLIOGRÁFICAS

[1] I. E. M. Rodríguez, L. M. T. Salazar, y D. F. C. Cámara, «Artificial intelligence as a support for non-pharmacological interventions against covid-19», *Rev. Peru. Med. Exp. Salud Publica*, vol. 37, n.º 3, pp. 582-584, 2020, doi: 10.17843/rpmesp.2020.373.5704.

[2] D. M. Espín Espín, S. M. Vilcacundo Pérez, y M. A. Banda Casa, «Aplicación de Machine Learning para el proceso de clasificación de datos asociados al centro médico Celina | Revista Científica y Tecnológica VICTEC», *VICTEC Rev. Académica Científica*, vol. 4, n.º 7, pp. 74-91, 2023.

[3] D. E. S. Alcalá *et al.*, «Clasificadores de inteligencia artificial basados en Radiómica del TC basal para predecir pronóstico funcional en pacientes con Hemorragia Intracerebral Espontánea (HIE).», *Seram*, vol. 1, n.º 1, Art. n.º 1, may 2022, Accedido: 23 de noviembre de 2022. [En línea]. Disponible en: https://www.piper.espacio-seram.com/index.php/seram/article/view/8440

[4] NACIONES-UNIDAS, *La Agenda 2030 y los Objetivos de Desarrollo Sostenible: una oportunidad para América Latina y el Caribe*, NACIONES UNIDAS. en 7MO CEPAL. Santiago - Chile: NACIONES UNIDAS, 2018.

[5] C. Ingino, «Innovación E Inteligencia Artificial En Medicina», Diagnostico Journal. [En línea]. Disponible en: https://philarchive.org/archive/INGIEI-2

[6] Moran Chaguay, Hermes Mick, «"Aplicación de Técnicas de Inteligencia Artificial Para el Proyecto Piloto de Telemedicina – Cayapas», Título profesional, UNIVERSIDAD DE GUAYAQUIL, Ecuador, Guayaquil-Ecuador, 2022.

[7] R. W. Bernaldo Joaquin, «Propuesta de implementación de la tecnología emergente basada en inteligencia artificial (AI) en diagnóstico médico en Hospital Cayetano Heredia».

[8] J. G. Corvalán, «Inteligencia artificial para la recuperación pospandemia. Informe 3», CAF, workingPaper, jun. 2022. Accedido: 23 de noviembre de 2022. [En línea]. Disponible en: https://cafscioteca.azurewebsites.net/handle/123456789/1922

[9] Aquino Arcata, Rene, Cuevas Machaca, Ronald, Godoy Montoya, Luis, y Rodríguez Puma, Heber, «Aplicación de regresión logística para la predicción de demanda por especialidad médica en consulta externa hospitalaria | Innovación y Software». Accedido: 23 de noviembre de 2022. [En línea]. Disponible en: https://revistas.ulasalle.edu.pe/innosoft/article/view/45

[10] J. A. . Gomez, J. M. Perez, J. Muñoz, y E. Alba, «Un Modelo para la Prediccion de Recidiva de Pacientes Operados de Cancer de Mama (CMO) Basado en Redes Neuronales», *Intel. Artif.*, vol. 4, n.º 11, p. 289, nov. 2000, doi: 10.4114/ia.v4i11.689.

[11] C. C. D. Martínez, I. B. Cruz, y J. E. C. Cogollo, «MODELO PREDICTIVO PARA EL PRONÓSTICO DE TIEMPOS DE ESTANCIA DE PACIENTES EN UNIDADES DE CUIDADOS INTENSIVOS», p. 29.

[12] F. Loncaric, O. Camara, G. Piella, y B. Bijnens, «La integración de la inteligencia artificial en el abordaje clínico del paciente: enfoque en la imagen cardiaca», *Rev. Esp. Cardiol.*, vol. 74, n.º 1, pp. 72-80, ene. 2021, doi: 10.1016/j.recesp.2020.07.012.

[13] E. F. D. Sullcahuamán, F. R. P. Pantigoso, E. S. Aguilar, y J. D. Sullcahuamán, «Inteligencia artificial y gobierno digital durante la covid 19 en una institución prestadora de salud, 2021-2022», *Cienc. Lat. Rev. Científica Multidiscip.*, vol. 6, n.º 4, Art. n.º 4, sep. 2022, doi: 10.37811/cl_rcm.v6i4.2896.

[14] M. Á. Vega, L. M. Q. Mora, y M. V. C. Badilla, «Inteligencia artificial y aprendizaje automático en medicina», *Rev. Medica Sinerg.*, vol. 5, n.º 8, Art. n.º 8, ago. 2020, doi: 10.31434/rms.v5i8.557.

[15] F. L. Sención Ramírez y E. Montero Beltré, «Centralización de historia médica electrónica utilizando inteligencia artificial: sistema ReCore», Thesis, Santo Domingo: Universidad Iberoamericana (UNIBE), 2021. Accedido: 19 de noviembre de 2022. [En línea]. Disponible en: https://repositorio.unibe.edu.do/jspui/handle/123456789/796

[16] V. E. Carrasco Chanamé, C. H. Cépeda Rodríguez, J. A. Rodríguez Durand, y R. Salinas Méndez, «Implementación de solución tecnológica en una empresa de asistencia de extensión de garantía, utilizando la inteligencia artificial». Universidad Peruana de Ciencias Aplicadas (UPC), 2019. [En línea]. Disponible en: http://hdl.handle.net/10757/625868

[17] Aira Céspedes, Francisco Antonio, Casas Moya, Luis Miguel, y Romero Naupari, Pablo Jesús, «Aplicación y casos de uso de técnicas de inteligencia artificial contra el COVID 19», *Rev. Investig. Sist. E Informática*, vol. 14, n.º 1, Art. n.º 1, dic. 2021, doi: 10.15381/risi.v14i1.21862.

[18] M. G. Sánchez Trujillo y J. Á. Pérez Hernández, *Metodología CRISP-DM en la gestión de proyecto de Data Mining. Caso enfermedades dermatológicas*. 2023. Accedido: 31 de octubre de 2023. [En línea]. Disponible en: https://repository.universidadean.edu.co/handle/10882/13087

[19] WEKA, «Weka 3 - Data Mining with Open Source Machine Learning Software in Java». Accedido: 3 de noviembre de 2023. [En línea]. Disponible en: https://www.cs.waikato.ac.nz/ml/weka/

[20] WIKIPEDIA, «Orange (software)», *Wikipedia*. 17 de octubre de 2023. Accedido: 3 de noviembre de 2023. [En línea]. Disponible en: https://en.wikipedia.org/w/index.php?title=Orange_(software)&oldid=1180592954

[21] M. R. M. Rodríguez, A. G. L. Montaño, L. R. Páez, R. R. Molina, y D. G. Cruz, «Aplicación de minería de datos mediante la metodología CRISP-DM para defunciones accidentales y violentas en Nezahualcóyotl», *Rev. Aristas*, vol. 10, n.º 18, Art. n.º 18, may 2023.

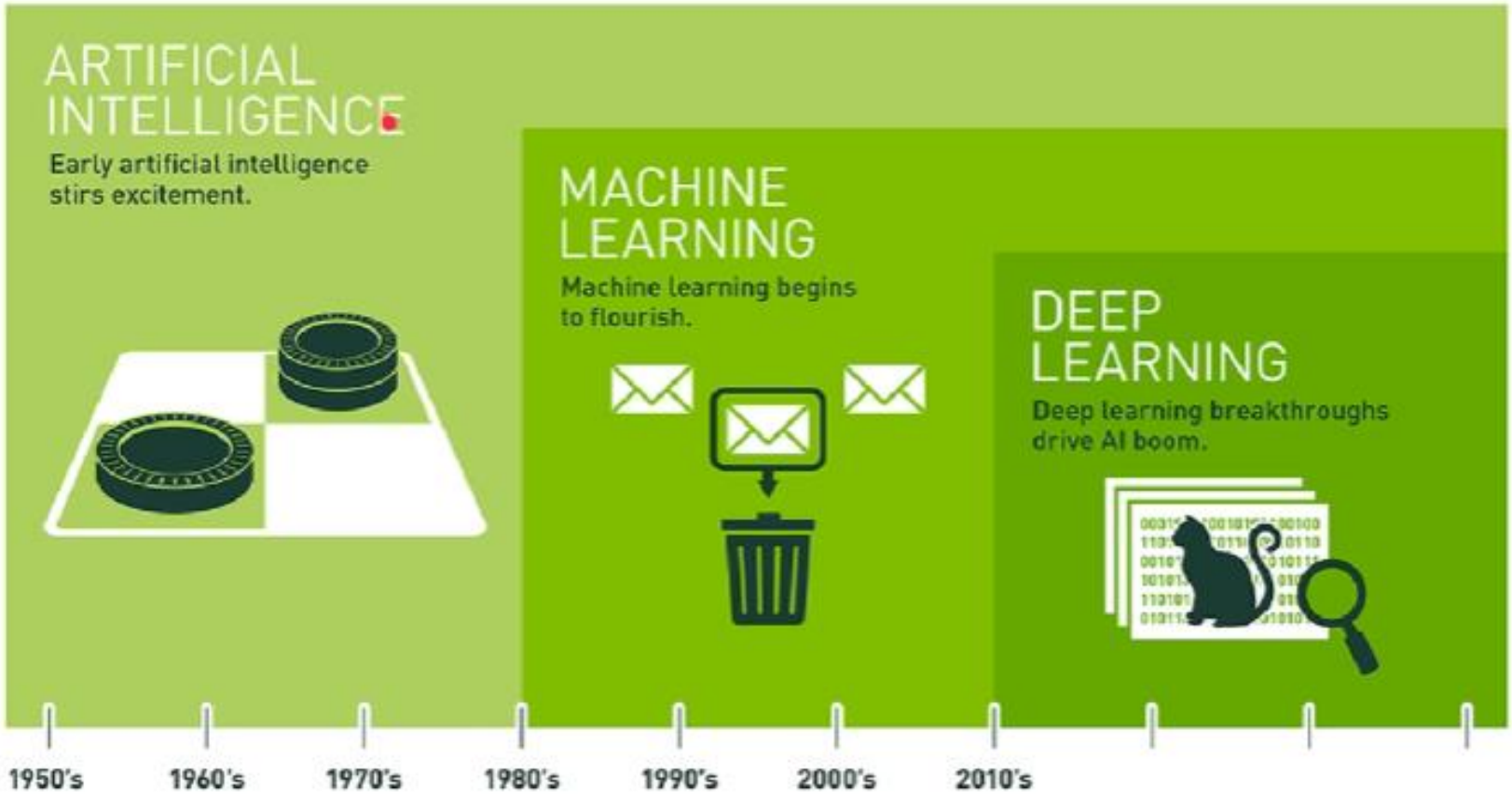
[22] I. S. Rivas Hidalgo, «Evaluación y aplicación de algoritmos de inteligencia artificial explicada para apoyar la toma de decisiones médicas en la salud fetal.», bachelorThesis, Quito : EPN, 2022, 2022. Accedido: 31 de octubre de 2023. [En línea]. Disponible en: http://bibdigital.epn.edu.ec/handle/15000/22551

[23] M. Garrido, «Machine Learning y Data Science con Python», Udemy. Accedido: 19 de enero de 2023. [En línea]. Disponible en: https://www.udemy.com/course/aprende-data-science-y-machine-learning-con-python/

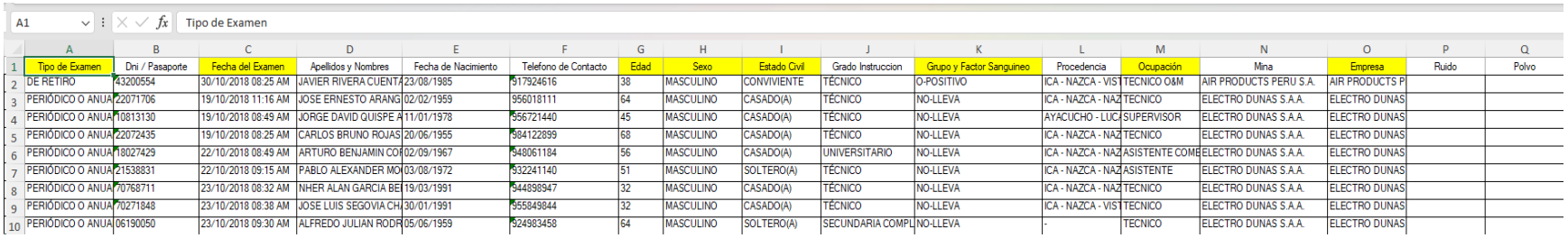
VIII. ANEXOS

Anexo 01: Objetivos del desarrollo sostenible de las Naciones Unidas [4]

Anexo 02: Evolución de la Inteligencia Artificial [23]



Anexo 03: Instrumento de recojo de datos



Tipo de Examen	Dni / Pasaporte	Fecha del Examen	Apellidos y Nombres	Fecha de Nacimiento	Telefono de Contacto	Edad	Sexo	Estado Civil	Grado Instruccion	Grupo y Factor Sanguineo	Procedencia	Ocupación	Mina	Empresa	Ruido	Polvo
DE RETIRO	43200554	30/10/2018 08:25 AM	JAVIER RIVERA CUENTA	23/08/1985	917924616	38	MASCULINO	CONVIVIENTE	TÉCNICO	O-POSITIVO	ICA - NAZCA - VIST	TECNICO O&M	AIR PRODUCTS PERU S.A.	AIR PRODUCTS P		
PERIÓDICO O ANUA	22071706	19/10/2018 11:16 AM	JOSE ERNESTO ARANG	02/02/1959	956018111	64	MASCULINO	CASADO(A)	TÉCNICO	NO-LLEVA	ICA - NAZCA - NAZ	TECNICO	ELECTRO DUNAS S.A.A.	ELECTRO DUNAS		
PERIÓDICO O ANUA	10813130	19/10/2018 08:49 AM	JORGE DAVID QUISPE A	11/01/1978	956721440	45	MASCULINO	CASADO(A)	TÉCNICO	NO-LLEVA	AYACUCHO - LUCA	SUPERVISOR	ELECTRO DUNAS S.A.A.	ELECTRO DUNAS		
PERIÓDICO O ANUA	22072435	19/10/2018 08:25 AM	CARLOS BRUNO ROJAS	20/06/1955	984122899	68	MASCULINO	CASADO(A)	TÉCNICO	NO-LLEVA	ICA - NAZCA - NAZ	TECNICO	ELECTRO DUNAS S.A.A.	ELECTRO DUNAS		
PERIÓDICO O ANUA	18027429	22/10/2018 08:49 AM	ARTURO BENJAMIN COI	02/09/1967	948061184	56	MASCULINO	CASADO(A)	UNIVERSITARIO	NO-LLEVA	ICA - NAZCA - NAZ	ASISTENTE COME	ELECTRO DUNAS S.A.A.	ELECTRO DUNAS		
PERIÓDICO O ANUA	21538831	22/10/2018 09:15 AM	PABLO ALEXANDER MO	03/08/1972	932241140	51	MASCULINO	SOLTERO(A)	TÉCNICO	NO-LLEVA	ICA - NAZCA - NAZ	ASISTENTE	ELECTRO DUNAS S.A.A.	ELECTRO DUNAS		
PERIÓDICO O ANUA	70768711	23/10/2018 08:32 AM	NHER ALAN GARCIA BE	19/03/1991	944898947	32	MASCULINO	CASADO(A)	TÉCNICO	NO-LLEVA	ICA - NAZCA - NAZ	TECNICO	ELECTRO DUNAS S.A.A.	ELECTRO DUNAS		
PERIÓDICO O ANUA	70271848	23/10/2018 08:38 AM	JOSE LUIS SEGOVIA CH	30/01/1991	955849844	32	MASCULINO	CASADO(A)	TÉCNICO	NO-LLEVA	ICA - NAZCA - VIST	TECNICO	ELECTRO DUNAS S.A.A.	ELECTRO DUNAS		
PERIÓDICO O ANUA	06190050	23/10/2018 09:30 AM	ALFREDO JULIAN RODR	05/06/1959	924983458	64	MASCULINO	SOLTERO(A)	SECUNDARIA COMPL	NO-LLEVA	-	TECNICO	ELECTRO DUNAS S.A.A.	ELECTRO DUNAS		

VigSegmentaria	VIBTotal	Cancerigenos	Mutagenicos	Solventes	MetalPesado	Temperaturas	Biologicos	Posturas	Turnos	Cargas	MovRepet	PantallaPVD	Otros	Describir	Antecedentes Personales
															ANTECEDENTES / OTROS PI
															ENFERMEDADES OTROS / D
															ANTECEDENTES / OTROS PI
															ANTECEDENTES / OTROS PI

Antecedentes Familiares	Habitos Nocivos Alcohol	Habitos Nocivos Tabaco	Habitos Nocivos Drogas	Numero Hijos	Alergias	Inmunizacion Influenza	Inmunizacion HepatitisB	Inmunizacion Tetano	Inmunizacion Fiebre Amarilla	Peso	Talla	IMC
PADRE / FALLECIDO, MADR	POCO	NO	NO	1	NO					72.0	1.63	27.10
PADRE / FALLECIDO, MADR	POCO	NO	NO	2	NO					93.2	1.68	33.02
PADRE / FALLECIDO, MADR	POCO	NO	NO	3	NO					70.0	1.57	28.40
PADRE / FALLECIDO, MADR	NO	NO	NO	4	NO					59.0	1.65	21.67
PADRE / NO REFIERE ANTE	POCO	NO	NO	3	NO					64.2	1.60	25.08
PADRE / NO REFIERE ANTE	NO	POCO	NO	0	NO					86.5	1.65	31.77
PADRE / HIPERTENSIÓN AR	POCO	POCO	NO	1	NO					101.0	1.74	33.36
PADRE / FALLECIDO, MADR	POCO	NO	NO	2	NO					78.0	1.68	27.64
PADRE / FALLECIDO, MADR	POCO	NO	NO	3	NO					83.0	1.71	28.38

AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG
Indice Cintura	Temperatura	Perimetro Cadera	Perimetro Abdominal	Presion Sistolica	Presion Diastolica	Frecuencia Respiratoria	Frecuencia Cardiaca	Saturacion Oxigeno	CabezaDescripcion	CuelloDescripcion	NarizDescripcion	BocaAdmigdalaFaringeLaringeDescripcion
0.9	36.5	102	91	100	70	18	81	97				
1.0	36.5	113	109	100	70	18	84	95				
0.9	36.5	99	88	110	70	18	55	96				
0.9	36.5	96	91	110	60	18	57	95				
0.9	37.1	96	88	100	70	20	84	98				
0.9	37.2	108	100	110	80	21	75	97				
0.9	36.5	119	105	110	70	18	81	98				
0.9	36.5	109	98	110	80	18	83	98				
0.9	36.5	109	102	120	70	18	78	98				

BH	BI	BJ	BK	BL	BM	BN	BO	BP	BQ
MiembrosSuperioresDescripcion	MiembrosInferioresDescripcion	ReflejosOsteoTendinososDescripcion	ColumnaDescripcion	MarchaDescripcion	AbdomenDescripcion	DescripcionGeneral	AnillosInguinalesDescripcion	HerniasDescripcion	VaricesDescripcion

BR	BS	BT	BU	BV	BW	BX	BY	BZ	CA	CB	CC	CD	CE
OrganosGenitalesDescripcion	GangliosDescripcion	LenAtenMemoOrientIntelAfect	Gingivitis	PiezasMalEstado	PiezasFaltan	T-PMalEstado	T-PFaltan	AntecedentesLaborales	AntecedentesPatologicos	VisionCercaScod	VisionCercaScoi	VisionCercaCcod	VisonCercaCcoi
				0	0	-	-			20/20	20/20	NO APLICA	NO APLICA
				0	0	-	-			20/100	20/100	20/25	20/25
				0	0	-	-			20/20	20/25	NO APLICA	NO APLICA
				0	0	-	-			20/100	20/100	20/30	20/30
				0	0	-	-			20/200	20/200	20/20	20/25
				0	0	-	-			20/20	20/20	NO APLICA	NO APLICA
				0	0	-	-			20/20	20/20	NO APLICA	NO APLICA
				0	0	-	-			20/20	20/20	NO APLICA	NO APLICA
				0	0	-	-			20/40	20/40	20/20	20/20

CF	CG	CH	CI	CJ	CK	CL	CM	CN	CO	CP	CQ	CR	CS	CT
VisionLejosScod	VisionLejosScoi	VisionLejosCcod	VisonLejosCcoi	VisonColores	EnfermedadesOculares	ReflejosPupilares	Conclusiones Oftalmología	Calibracion	Otoscopia OidoDerecho	Otoscopia OidoIzquierdo	VAOD125	VAOD250	VAOD500	VAOD1000
20/20	20/20	NO APLICA	NO APLICA				EMÉTROPE	JUNIO 2018	CAE permeable, Membrana Ti	CAE permeable, Membrana Tir			15	10
20/20	20/20	NO APLICA	NO APLICA				AMETROPÍA CORREGIDA	JUNIO 2018	CAE permeable, Membrana Ti	CAE permeable, Membrana Tir			5	10
20/20	20/25	NO APLICA	NO APLICA				EMÉTROPE	JUNIO 2018	CAE permeable, Membrana Ti	CAE permeable, Membrana Tir			15	5
20/30	20/30	NO APLICA	NO APLICA				AMETROPÍA NO CORREGIDA	JUNIO 2018	CAE permeable, Membrana Ti	CAE permeable, Membrana Tir			25	15
20/40	20/100	NO APLICA	NO APLICA				AMETROPIA	JUNIO 2018	CAE permeable, Membrana Ti	CAE permeable, Membrana Tir			10	5
20/25	20/25	NO APLICA	NO APLICA				EMÉTROPE	JUNIO 2018	CAE permeable, Membrana Ti	CAE permeable, Membrana Tir			10	5
20/20	20/20	NO APLICA	NO APLICA				EMÉTROPE	JUNIO 2018	CAE permeable, Membrana Ti	CAE permeable, Membrana Tir			15	5
20/20	20/25	NO APLICA	NO APLICA				EMÉTROPE	JUNIO 2018	CAE permeable, Membrana Ti	CAE permeable, Membrana Tir			15	10
20/30	20/30	20/20	20/20				AMETROPÍA CORREGIDA	JUNIO 2018	CAE permeable, Membrana Ti	CAE permeable, Membrana Tir			5	15

CU	CV	CW	CX	CY	CZ	DA	DB	DC	DD	DE	DF	DG	DH	DI	DJ	DK	DL
VAOD2000	VAOD3000	VAOD4000	VAOD6000	VAOD8000	VAOI125	VAOI250	VAOI500	VAOI1000	VAOIi2000	VAOI3000	VAOI4000	VAOI6000	VAOI8000	VOOD125	VOOD250	VOOD500	VOOD1000
10	20	10	5	15			15	10	5	15	15	10	10				
15	10	15	10	15			15	5	5	10	25	10	15				
15	10	15	15	10			10	15	25	25	25	15	20				
5	5	10	10	20			25	5	10	10	15	15	25				
5	15	25	10	25			15	5	10	25	20	25	25				
10	15	25	15	25			15	5	5	15	15	10	25				
5	10	10	15	10			10	5	0	20	20	10	15				
10	10	5	5	0			15	10	10	5	5	0	10				
20	15	20	10	25			5	15	20	15	25	20	20				

DM	DN	DO	DP	DQ	DR	DS	DT	DU	DV	DW	DX	DY	DZ	EA	EB	EC	ED
VOOD2000	VOOD3000	VOOD4000	VOOD6000	VOOD8000	VOOI125	VOOI250	VOOI500	VOOI1000	VOOI2000	VOOI3000	VOOI4000	VOOI6000	VOOI8000	EM-A-OD125	EM-A-OD 250	EM-A-OD 500	EM-A-OD1000

EE	EF	EG	EH	EI	EJ	EK	EL	EM	EN	EO	EP	EQ	ER	ES	ET	EU
EM-A-OD 2000	EM-A-OD 3000	EM-A-OD 4000	EM-A-OD 6000	EM-A-OD 8000	EM-A-OI125	EM-A-OI 250	EM-A-OI 500	EM-A-OI 1000	EM-A-OI 2000	EM-A-OI 3000	EM-A-OI 4000	EM-A-OI 6000	EM-A-OI 8000	EM-O-OD 125	EM-O-OD 250	EM-O-OD 500
														2		

EV	EW	EX	EY	EZ	FA	FB	FC	FD	FE	FF	FG	FH	FI	FJ	FK	FL
EM-O-OD 1000	EM-O-OD 2000	EM-O-OD 3000	EM-O-OD 4000	EM-O-OD 6000	EM-O-OD 8000	EM-O-OI 125	EM-O-OI 250	EM-O-OI i500	EM-O-OI 1000	EM-O-OI 2000	EM-O-OI 3000	EM-O-OI 4000	EM-O-OI 6000	EM-O-OI 8000	Conclusiones Audiometría	FVC
															NORMOACUSIA BILATERAL	4.35
															NORMOACUSIA BILATERAL	
															NORMOACUSIA BILATERAL	
															NORMOACUSIA BILATERAL	
															NORMOACUSIA BILATERAL	
															NORMOACUSIA BILATERAL	
															NORMOACUSIA BILATERAL	
															NORMOACUSIA BILATERAL	
															NORMOACUSIA BILATERAL	

FM	FN	FO	FP	FQ	FR	FS	FT	FU	FV	FW	FX	FY	FZ	GA	GB
VEF1	VEF CVF	Conclusion Espirometría	0/-	0/0	0/1	1/0	1/1	1/2	2/1	2/2	2/3	3/1	3/2	3/+	Conclusion RayosX
3.60	82.8	FUNCIÓN VENTILATORIA N(	-	X	-	-	-	-	-	-	-	-	-	-	RADIOGRAFÍA DE PULMONES NORMAL
			-	X	-	-	-	-	-	-	-	-	-	-	RADIOGRAFÍA DE PULMONES NORMAL
			-	X	-	-	-	-	-	-	-	-	-	-	RADIOGRAFÍA DE PULMONES NORMAL
			-	X	-	-	-	-	-	-	-	-	-	-	RADIOGRAFÍA DE PULMONES NORMAL
			-	X	-	-	-	-	-	-	-	-	-	-	RADIOGRAFÍA DE PULMONES NORMAL
			-	X	-	-	-	-	-	-	-	-	-	-	RADIOGRAFÍA DE PULMONES NORMAL
			-	X	-	-	-	-	-	-	-	-	-	-	RADIOGRAFÍA DE PULMONES NORMAL
			-	X	-	-	-	-	-	-	-	-	-	-	RADIOGRAFÍA DE PULMONES NORMAL
			-	X	-	-	-	-	-	-	-	-	-	-	ENFISEMA, NO ESPECIFICADO

GC	GD	GE	GF	GG	GH	GI	GJ	GK	GL	GM	GN	GO
RitmoCardiaco	Frecuencia Cardiaca	EjeCardiaco	IntervaloPR	ComplejoQRS	IntervaloQTC	DxEKG	Hemoglobina	Hematocrito	NeutrofilosSegementados	Abastonados	Segmentados	Monocitos
							16.40	50.70	49	0	0	7
						EKG NO PATOLOG	15.70	48.60	52	0	0	5
						EKG NO PATOLOG	16.80	50.30	46	0	0	7
						EKG NO PATOLOG	14.80	47.10	51	0	0	10
						RESULTADOS AN(	13.90	42.60	70	0	0	6
						EKG NO PATOLOG	16.20	49.40	51	0	0	7
							15.90	48.30	55	0	0	7
							15.50	48.20	45	0	0	13
						EKG NO PATOLOG	15.90	48.70	55	0	0	8

GP	GQ	GR	GS	GT	GU	GV	GW	GX	GY	GZ	HA	HB	HC	HD	HE	HF
Eosinofilos	Basofilos	Linfocitos	Hematies	Plaquetas	Mielocitos	Juveniles	Glucosa	ColesterolTotal	ColesterolHDL	ColesterolLDL	VLDL	Trigliceridos	TGO	TGP	CreatininaSuero	Color
4	1	39	5.320 000	254,000	0	0	100.0	179.0	41.8	122.5	122.5	82.0			0.97	AMARILLO CLARO
2	1	36	5.020 000	164,000	0	0	88.0	268.0				179.0			0.84	AMARILLO CLARO
1	1	45	5.030 000	247,000	0	0	85.0	179.0				153.0			0.96	AMARILLO CLARO
3	1	35	4.870 000	210,000	0	0	91.0	225.0				171.0			1.09	AMARILLO CLARO
2	1	21	4.440 000	261,000	0	0	103.0	189.0				166.0			1.22	AMARILLO CLARO
7	1	34	5.170 000	258,000	0	0	94.0	251.0				204.0			0.95	AMARILLO CLARO
1	1	36	5.25	294	0	0	93.0	158.0				111.0			1.06	AMARILLO CLARO
2	1	40	5.22	219	0	0	107.0	156.0				139.0			0.98	AMARILLO CLARO
2	0	25	5.110 000	250,000	0	0	105.0	223.0				183.0			0.89	AMARILLO CLARO

HG	HH	HI	HJ	HK	HL	HM	HN	HO	HP	HQ	HR	HS	HT	HU	HV
Aspecto	Densidad	ReaccionPH	GlucosaExamenCompletoOrina	Bilirrubinas	Cetonas	ProteinasTotales	Urobilinogeno	Leucocitos	Hematies_	Levaduras	CristDeoxalatosCalcio	Germenes	Albumina	LeucocitosTotales	Piocitos
LIGERAMENTE TURBIO	1.025	6.0	NEGATIVO											6.380	
TRANSPARENTE	1.005	7.0	NEGATIVO											5.470	
LIGERAMENTE TURBIO	1.015	5.0	NEGATIVO											7.770	
LIGERAMENTE TURBIO	1.025	5.0	NEGATIVO											5.450	
TRANSPARENTE	1.020	5.0	NEGATIVO											7.440	
TRANSPARENTE	1.015	5.0	NEGATIVO											5.240	
TRANSPARENTE	1.025	5.0	NEGATIVO											9.47	
TRANSPARENTE	1.030	5.0	NEGATIVO											8.61	
LIGERAMENTE TURBIO	1.020	5.0	NEGATIVO											8.790	

HW	HX	HY	HZ	IA	IB	IC	ID	IE	IF	IG	IH	II	IJ
FilamentosMucoides	CCVolumenCorpuscularMedio	CCHBCorpuscular	LuesVDRL	PlomoSangre	Magnesio	Cocaina	Marihuana	Conclusiones Psicosensométrico	CoeficienteIntelectual	AptitudPsicologica	Conclusiones Psicología	Recomendaciones	Conclusiones Odontología
	95.40	30.80									ESTRÉS MEDIO/ EVALUACIÓN PSICOLÓGICA NOR		
	96.80	31.40									EVALUACIÓN PSICOLÓGICA NORMAL		
	100.00	33.40									EVALUACIÓN PSICOLÓGICA NORMAL		
	96.70	30.40									EVALUACIÓN PSICOLÓGICA NORMAL		
	95.90	31.30									EVALUACIÓN PSICOLÓGICA NORMAL		
	95.40	31.30									EVALUACIÓN PSICOLÓGICA NORMAL		
	92.00	30.03									EVALUACIÓN PSICOLÓGICA NORMAL		
	92.03	29.08									EVALUACIÓN PSICOLÓGICA NORMAL		

IK	IL	IM	IN	IO	IP	IQ	IR	IS	IT
Conclusiones Triaje	Conclusiones Medicina	ClinicaRealizoEmo	ConclusionLabo	Dx-T	RecomendacionesConcatenadas	RestriccionConcatenadas	AptitudFinal	MotivoObservacion	VigenciaEmo
SOBREPESO	EVALUACIÓN OSTEOMUSCULAR SIN ALTERACIÓ		RESULTADOS DE LABORAT	EXAMEN FÍSICO S	CHEQUEO MÉDICO PREVENTIVO AN		APTO		30/10/2019
OBESIDAD TIPO I	EVALUACIÓN OSTEOMUSCULAR SIN ALTERACIÓ		HIPERLIPIDEMIA MIXTA	HIPERLIPIDEMIA M	LA OBESIDAD ES UN FACTOR DE RI		APTO		23/10/2019
SOBREPESO	EVALUACIÓN OSTEOMUSCULAR SIN ALTERACIÓ		RESULTADOS DE LABORAT	SOBREPESO	SE LE RECOMIENDA REDUCIR SU P		APTO		23/10/2019
NORMOPESO	EVALUACIÓN OSTEOMUSCULAR SIN ALTERACIÓ		HIPERLIPIDEMIA MIXTA	HIPERLIPIDEMIA M	EVALUACION POR MEDICINA INTER		APTO		19/10/2019
SOBREPESO	EVALUACIÓN OSTEOMUSCULAR SIN ALTERACIÓ		RESULTADOS DE LABORAT	SOBREPESO, AME	ACUDIR A OFTALMOLOGIA PARA US		APTO		22/10/2019
OBESIDAD TIPO I	EVALUACIÓN OSTEOMUSCULAR SIN ALTERACIÓ		HIPERLIPIDEMIA MIXTA	HIPERLIPIDEMIA M	LA OBESIDAD ES UN FACTOR DE RI		APTO		22/10/2019
OBESIDAD TIPO I	EVALUACIÓN OSTEOMUSCULAR SIN ALTERACIÓ		RESULTADOS DE LABORAT	OBESIDAD TIPO I	LA OBESIDAD ES UN FACTOR DE RI		APTO		23/10/2019
SOBREPESO	EVALUACIÓN OSTEOMUSCULAR SIN ALTERACIÓ		RESULTADOS DE LABORAT	SOBREPESO	SE LE RECOMIENDA REDUCIR SU P		APTO		23/10/2019
SOBREPESO	EVALUACIÓN OSTEOMUSCULAR SIN ALTERACIÓ		HIPERCOLESTEROLEMIA/ P	HIPERCOLESTER(	SE LE RECOMIENDA REDUCIR SU P		APTO		24/10/2019

Anexo 04: Empresas de los trabajadores evaluados por sexo

	Sexo		
Count	FEMENINO	MASCULINO	Total
AGENCIA DE SEGURIDAD PRIVADA ESCORPION SOCIEDAD COMERCIAL DE RESPONSABILIDAD LIMITADA	0.0	2.0	2.0
AIR PRODUCTS PERU S.A.	0.0	3.0	3.0
ANDROMEDA PROYECTOS Y CONSTRUCCIONES SOCIEDAD COMERCIAL DE RESPONSABILIDAD LIMITADA	0.0	1.0	1.0
ANGELES MINERIA Y CONSTRUCCION S.A.C.	2.0	15.0	17.0
APC CORPORACION S.A.	71.0	99.0	170.0
APM TERMINALS INLAND SERVICES S.A.	0.0	2.0	2.0
AUSENCO PERU S.A.C.	0.0	1.0	1.0
AUTORIDAD NACIONAL DEL AGUA	5.0	18.0	23.0
AZAÑEDO QUILCATE INGENIEROS CONTRATISTAS S.R.L.	0.0	17.0	17.0
CABALLERO CONTRATISTAS MINEROS SOCIEDAD ANONIMA CERRADA - CABAMI S.A.C.	0.0	5.0	5.0
CANTALLOC SOCIEDAD COMERCIAL DE RESPONSABILIDAD LIMITADA - CANTALLOC S.R.L.	2.0	36.0	38.0
CENTURY ECOLOGICAL CORPORATION S.A.C. - ECOCENTURY S.A.C.	0.0	7.0	7.0
CLINICA INTERNACIONAL S A	0.0	2.0	2.0
COMIN S.A.C.	0.0	13.0	13.0
CONSTRUCTORA ALEXANDER E.I.R.L.	0.0	11.0	11.0
CONSULTING ARSI E.I.R.L.	0.0	2.0	2.0
CORPORACION GUERRERO & BAZALAR S.A.C.	1.0	3.0	4.0
DISTRIBUIDORA KERAMIK F & G E.I.R.L.	0.0	3.0	3.0
DISTRIBUIDORA LOGISTICA Y TRANSPORTES SA	0.0	1.0	1.0
ELECTRO DUNAS S.A.A.	2.0	15.0	17.0
EMAPAVIGS S.A.	7.0	27.0	34.0
EMPRESA DE BIENES Y SERVICIOS PARA EL HOGAR S.A.C.	1.0	1.0	2.0
EMPRESA DE SERVICIOS MULTIPLES J.M. INGENIEROS E.I.R.L.	0.0	9.0	9.0
EXPLOSUPPORT SAC	0.0	16.0	16.0
EXSA S A	1.0	1.0	2.0

Empresa				
	EXSA S A	1.0	1.0	2.0
	FEPALC E.I.R.L.	0.0	1.0	1.0
	FERREYROS SOCIEDAD ANÓNIMA	0.0	4.0	4.0
	GRUPO J Q SOCIEDAD ANONIMA CERRADA	0.0	6.0	6.0
	HASA CONSTRUCT SERVICIOS GENERALES S.A.C.	1.0	16.0	17.0
	HIMAPE CONTRATISTAS MINEROS SAC	0.0	18.0	18.0
	HLC INGENIERIA Y CONSTRUCCION SOCIEDAD ANONIMA CERRADA - HLC S.A.C.	0.0	8.0	8.0
	HUAYNA S.A.C.	0.0	2.0	2.0
	HV CONTRATISTAS S.A.	0.0	17.0	17.0
	INGESUR EMPRESA INDIVIDUAL DE RESPONSABILIDAD LIMITADA	0.0	7.0	7.0
	JJ GENERAL SERVICES SOCIEDAD ANONIMA CERRADA	0.0	3.0	3.0
	MALCOHUACCHA HUAMANI DANITZA	0.0	1.0	1.0
	MARCOBRE S.A.C.	0.0	2.0	2.0
	MEGAPACK TRADING SAC	0.0	3.0	3.0
	MINE SENSE SOLUTIONS SERVICES S.A.C.	0.0	2.0	2.0
	MINERA ARTESANAL SAN LUIS SOCIEDAD ANONIMA	0.0	16.0	16.0
	MINERA VIRGEN DE CHAPI 87 DE ICA SOCIEDAD ANONIMA CERRADA	0.0	4.0	4.0
	MINERIA CORPORATIVA SOCIEDAD ANONIMA CERRADA	0.0	1.0	1.0
	MINERIA Y EXPORTACIONES S.A.C.	2.0	17.0	19.0
	MOVIMIENTO Y LOGISTICA S.R.L.	0.0	2.0	2.0
	NESSUS HOTELES PERU S.A.	4.0	10.0	14.0
	OMEGANET PERU S.A.C. / POLICLINICO JUAN PABLO ELIAS SCRL	1.0	0.0	1.0
	OPERACIONES SERVICIOS Y SISTEMAS S.R.L.	0.0	6.0	6.0
	PEGASUS RENT A CAR S.A.C.	0.0	8.0	8.0
	PROINGCOM S.A.C.	0.0	3.0	3.0
	RED OPERADORA DE ENERGIA SAC	3.0	3.0	6.0
	SANTO DOMINGO CONTRATISTAS GENERALES S.A	0.0	1.0	1.0
	SANTO DOMINGO CONTRATISTAS GENERALES S.A	0.0	1.0	1.0
	SERVICIOS GENERALES TRANSMARR S.A.C	0.0	2.0	2.0
	SOLUCIONES INTEGRALES JYDS E.I.R.L.	0.0	7.0	7.0
	STEO INVERSIONES Y SERVICIOS GENERALES S.A.C.	0.0	1.0	1.0
	TERRAMOVE S.A.C	10.0	59.0	69.0
	TRANSPORTES HAGEMSA S.A.C.	0.0	6.0	6.0
	VIDITEK S.A.C.	0.0	2.0	2.0
	VYT CONTRATISTAS SOCIEDAD ANONIMA CERRADA - VYT CONTRATISTAS S.A.C.	0.0	22.0	22.0
	ZONA REGISTRAL NRO XI SEDE ICA	6.0	5.0	11.0
	Total	119.0	575.0	694.0

Enmarcado con color rojo se resaltan las empresas que mayor cantidad de trabajadores han sido evaluados.