



Diseño de un prototipo para la generación de alarmas a partir del análisis de tendencias del perfil glucémico de un grupo de pacientes con diabetes

Anderson Mosquera Ruiz, Maira García Jaramillo

**Universidad EAN
Bogotá, Colombia**

Resumen

La diabetes es una enfermedad metabólica crónica caracterizada por elevados niveles de glucosa en plasma debido a que quien la padece no produce insulina (diabetes tipo 1) o la produce, pero hay una resistencia a la utilización de esta (diabetes tipo 2). El monitoreo continuo de los niveles de glucosa permite obtener un registro de los valores de este cada cierto tiempo, para que a partir de dicha información el personal clínico determine la efectividad del tratamiento. En este proceso es fundamental analizar posibles valores anómalos sobre los que el personal médico puede determinar estrategias para optimizar la terapia, que incluye fármacos y dieta, y que de esta manera la cantidad de valores anómalos de glucosa disminuya. En este trabajo se realiza una revisión bibliográfica de los métodos de determinación de anomalías en un set de datos, se realiza un análisis exploratorio de datos sobre datos de monitorización continua de glucosa de diferentes pacientes para determinar las métricas y variables a usar en análisis de las tendencias. Finalmente, se realiza un análisis con técnicas de machine learning para determinar tendencias e identificar valores anómalos en los perfiles glucémicos recopilados.

Palabras clave: detección de anomalías; anomalías; series de tiempo; perfil glucémico

Abstract

Diabetes is a chronic metabolic disease characterized by high plasma glucose levels because the sufferer does not produce insulin (type 1 diabetes) or produces it but there is a resistance to the use

of it (type 2 diabetes). The continuous monitoring of glucose levels allows obtaining a record of the values of the glucose in defined intervals, so that from this information the clinical staff determines the effectiveness of the treatment. In this process it is essential to analyze possible anomalous values on which the medical staff can determine strategies to optimize therapy, which includes drugs and diet, and that in this way decreasing the amount of anomalous glucose values. In this work, a bibliographic review of the methods of determining anomalies in a data set is carried out, an exploratory analysis of data on continuous glucose monitoring data from different patients is carried out to determine the metrics and variables to be used in the analysis of trends. Finally, an analysis with machine learning techniques is carried out to determine trends and identify anomalous values in the collected glycemetic profiles.

Keywords: anomaly detection; anomaly; time series; glycemetic profile

1. Introducción

De acuerdo con la Federación Internacional de Diabetes (IDF), el número de personas entre 20 y 79 años que padecen diabetes en el mundo es de 537 millones, lo que representa un gasto público asociado de USD 966 mil millones, es decir, un aumento del 316% en los últimos 15 años. Se estima que para el 2030 el número de personas con diabetes ascenderá a 643 millones y en el 2045 a 783 millones (IDF ATLAS 2021). Una cifra alarmante es la cantidad de muertes asociadas. En 2021 murieron 6.7 millones de personas, lo que equivale a 1 fallecimiento cada 5 segundos por esta enfermedad.

En el contexto colombiano, la prevalencia actual de la diabetes es del 9.9%, lo que representa cerca de 3.5 millones de personas (IDF ATLAS 2021), siendo el segundo país en Suramérica con mayor prevalencia de diabetes atribuible a problemas nutricionales y falta de actividad física. De acuerdo con el último informe del departamento administrativo nacional de estadística (DANE), la diabetes se sitúa entre las primeras cinco causas conocidas de muerte en Colombia, aumentado considerablemente en los últimos años. Adicionalmente, la población con diabetes fue una de las más afectadas durante la pandemia del coronavirus 2 del síndrome respiratorio agudo grave (SARS-CoV-2), más conocido como coronavirus disease 2019 (COVID-19), mostrando una prevalencia de muerte de 82% (1,82 veces) mayor que la de población sin esta enfermedad (Saha et al. 2021).

La evidencia acumulada en años recientes ha permitido demostrar que además de un diagnóstico temprano de la enfermedad, es necesario un adecuado control de la diabetes con el fin de reducir la progresión de complicaciones crónicas, y así mismo, la creciente carga financiera para los sistemas de atención de la salud ya saturados (Castillo et al. 2019). Para lograr el control glucémico, se requiere el uso de fármacos y en algunos casos terapias intensivas de insulina, dietas, ejercicio, junto a un automonitoreo con glucometría capilar y/o monitorización continua de glucosa. La monitorización de los niveles de glucosa se realiza a través de dispositivos médicos tales como glucómetros y sistemas de monitorización continua de glucosa, los primeros usan una tira radioactiva que requiere una gota de sangre para determinar el nivel de glucosa capilar. El sistema de monitoreo continuo requiere un sensor que dependiendo del fabricante puede tener una duración de 6



a 14 días, el sensor que se adhiere a la piel (antebrazo, abdomen, pierna) toma muestras de glucosa intersticial cada 5 minutos, los datos del sensor se visualizan en una pantalla y algunos de estos dispositivos cuentan con su propio software para descarga y visualización de los datos.

El paciente que cuenta con sistemas de monitorización continua de glucosa generalmente acude a su cita médica con el dispositivo y allí se realiza la descarga de datos y análisis por parte del especialista, lo que toma un tiempo considerable de la consulta. Por tanto, que el personal clínico pueda contar con herramientas que permitan identificar de forma ágil y fiable posibles tendencias que afecten negativamente el tratamiento del paciente resultaría en un aporte relevante para la toma de decisiones clínicas y mayor efectividad en el tratamiento.

En este artículo se realiza un análisis exploratorio para identificar posibles tendencias en los perfiles de glucosa de los pacientes que puedan afectar su tratamiento. Para ello primero se realiza una revisión de literatura de los métodos de determinación de anomalías en un set de datos, para identificar cuáles son las métricas y variables adecuadas para el conjunto de datos a explorar. Luego se procede a realizar un preprocesamiento de los datos de 10 pacientes, para la posterior aplicación de técnicas de machine learning que permitan determinar tendencias e identificar valores anómalos en los perfiles glucémicos recopilados.

2. Marco Teórico

Que son las anomalías en una serie de tiempo

Una serie de tiempo es una serie de observaciones que se han tomado de manera ordenada y se correlacionan con el tiempo (Blázquez-García, Conde, Mori, Lozano, 2021, p. 1). Algunas veces se pueden encontrar observaciones que son raras en el contexto particular de la serie de tiempo, estas se conocen como anomalías y se pueden definir como "observaciones que se desvían tanto de otras observaciones como para generar sospechas de que se han generadas por otros mecanismos" (Hilal, Gadsden, Yawney, 2021, p. 2) y su detección ha sido estudiada en diferentes áreas del conocimiento (Chandola, Banerjee, Kumar, 2009, p. 1), como el manejo de riesgos médicos o financieros, seguridad, aplicaciones de inteligencia artificial, entre otros (Pang, Shen, Cao, van den Hegel, 2021, p. 1).

Pang et al. (2021, p. 2) exponen que las anomalías pueden ser producidas por problemas en los sensores que capturan la información, por el ruido provocado por interferencias en la captura o transmisión de la información o la intromisión de datos que son erróneos o no deseados en la serie de tiempo; también se pueden producir por cambios en el comportamiento normal (Chandola et al., 2009, p. 3), lo que dificulta el trabajo de los analistas que buscan detectar y tratar estas anomalías.

Tipos de anomalías

Las anomalías puntuales son el caso más común de detección de anomalías en una serie de tiempo (Blázquez-García et al., 2021, p. 7), se presentan en series univariadas (De acuerdo con Collins (s.f. a) se pueden entender como series de tiempo en las cuales hay una sola variable dependiente



del tiempo) o multivariadas (series de tiempo en la cuales hay varias variables dependientes del tiempo que no necesariamente son independientes entre sí (Collins (s.f. b)) y para detectarlas se puede predecir el dato y establecer rangos por fuera de los cuales el punto observado sea una anomalía. También se pueden encontrar anomalías subsecuentes, las cuales son una colección de observaciones seguidas que se comportan de la manera descrita en la definición de anomalía (Blazquez-Garcia et al., 2021, p. 17), en estos casos se pueden establecer rangos de observaciones y compararlos entre sí para determinar cuáles son anómalos; las anomalías también se pueden presentar en series multivariadas como variables dependientes que se comportan de manera anómala (Blazquez-Garcia et al., 2021, p. 25), en este tipo de anomalías se considera toda la serie de tiempo y determina si existe la anomalía o no usando reducción de dimensionalidad o métodos de agrupamiento.

Técnicas de detección de anomalías

A continuación, se realizará una breve revisión de los tipos de métodos usados en la detección de anomalías con técnicas de machine learning:

- Técnicas basadas en distancia: Este tipo de método busca conocer la distancia entre la observación y las demás observaciones para determinar si es una observación anómala o no, la consideración es que si la distancia entre el punto y el grupo de puntos es muy grande la observación es anómala (Pang et al., 2021, p. 16). Ramaswamy, Rastogi, Shim (2000, p. 1) proponen una técnica que utiliza el método de K vecinos más cercanos para determinar cuántos puntos pueden ser considerados anómalos en un set de datos basándose en la distancia entre estos y el clúster de puntos; Pang, Ting, Albrecht (2015, p. 2) introducen el concepto de Vecino más cercano con menor parecido (LeSiNN: Least Similar Nearest Neighbors) en el cual se busca detectar los datos cuya similaridad con su vecino más cercano sea amplia, ya que si bien estos datos estarán alejados del grupo de puntos, pueden tener vecinos cercanos con los cuales no comparten características, por tanto pueden ser identificados como datos anómalos. Sugiyama y Borgwardt (2013, p. 3) proponen obtener una muestra aleatoria de los datos para determinar cuáles son anómalos y concluyen que haciéndolo de esta manera se pueden sobreponer retos de escalabilidad asociados con las técnicas de determinación de anomalías basadas en distancia aplicados en datasets de grandes dimensiones. Cuando se requiere analizar datos dispersos con patrones implícitos se puede utilizar el método de determinación de un Factor local de anomalía basado en distancia (LDOF por sus siglas en ingles) el cual determina la distancia relativa entre un punto y un grupo de puntos dispersos para luego asignarle a cada uno un factor de anomalía usado en la determinación de las anomalías (Zhang, Hutter, Jin, 2009, p. 814).

También se puede incluir en este tipo de métodos las medidas que se realizan sobre la agrupación de datos (clustering) (Pang et al., 2021, p. 19), como el factor de anomalía basado en clúster (CBLOF: cluster-Base outlier factor) que usa la distancia entre el punto y el centroide del clúster (He, Xu, Deng, 2002, p. 1644) o el método INN (Improved Nearest Neighbor) que además de verificar si el punto pertenece o no a un clúster, verifica la distancia entre el punto y otros clúster (Jiang, Song, Wang, Han, Li, 2006, p. 803). Jian, Tseng, Su (1999, p. 692) proponen un método denominado clusterización de dos fases en la cual obtienen sucesivos clústeres de tamaño reducido utilizando los datos residuales de clusterizaciones previas y mediante el diámetro del clúster y otros



parámetros determinan si los datos contenidos en este son o no anomalías. Una ventaja de este tipo de métodos es que hay abundante literatura al respecto, pero requieren bastante poder computacional (Pang et al., 2021, p. 17).

- **Medidas de clasificación:** Este tipo de métodos busca identificar una característica en el set de datos y definir si un nuevo dato que ingresa al set de datos tiene esa característica o no, si no la tiene el dato será considerado anómalo (Pang et al., 2021, p. 17). Cortés y Vapnik (1995, p. 278) explican que las máquinas de soporte vectorial pueden usarse como herramientas de clasificación supervisada y con ellas se puede identificar y clasificar en el set de datos las observaciones anómalas.
- **Modelos de ranking:** En este tipo de métodos se busca ordenar el set de datos usando el grado de anormalidad percibido en una variable dentro de las observaciones (Pang et al., 2021, p. 21), el problema es que se requiere que algunas anomalías vengan etiquetadas y de esa manera entrenar el modelo para hallar esa anomalía o tipo de anomalía particular dentro del set de datos (Pang et al., 2021, p. 23). También se puede presentar un caso extendido en el cual se alimentan al modelo no solo anomalías etiquetadas sino también clasificadas por el grado de anormalidad que presentan (Pang et al., 2021, p. 23), luego usando modelos de IRL (Inverse Reinforcement Learning) se puede determinar el modelo usado para generar esta clasificación (Ng, A y Russel S., 2000, p. 3) y de esa manera clasificar nuevas observaciones que se incorporen en el set de datos. Otro método que puede ser usado es la clasificación de las observaciones basándose en su probabilidad de ocurrencia en un set de datos, para esto se puede asumir que las observaciones normales tienen una alta probabilidad de ocurrencia, mientras que las anomalías son raras y por ende tienen una baja probabilidad de ocurrencia (Pang et al., 2021, p. 24).

Perfiles glucémicos

La diabetes es una enfermedad crónica, en la que el cuerpo no disuelve la glucosa de manera apropiada (Jain, Joshi, Mohanty, 2021, p. 1). Existen 3 tipos bien diferenciados de diabetes: (1) Diabetes Tipo 1, en la que el cuerpo no genera insulina para la regulación de glucosa; (2) Diabetes Tipo 2, en la que el cuerpo genera pequeñas cantidades de insulina que no son suficientes para la regulación de la glucosa y (3) Diabetes Gestacional, en la que el cuerpo de la madre gestante no produce insulina (Jain et al., 2021, p. 4). Un perfil glucémico es la representación gráfica de una serie de tiempo donde el eje y contiene los niveles de glucosa y el eje x contiene el tiempo; de estos perfiles se pueden realizar observaciones para optimizar los niveles de glucosa a diferentes horas o momentos del día (Bando, Ebe, Kato, Bando, Yonei, 2018, p. 1) dependiendo de la rutina del paciente y sus preferencias dietarias. Para obtener un perfil glucémico se realizan mediciones continuas mediante técnicas invasivas, semi-invasivas o no invasivas de los niveles de glucosa en la sangre; (Jain et al., 2021, p. 8), con lo cual se puede realizar un análisis de la serie temporal resultante para determinar el nivel de variabilidad de la glucosa en la sangre y determinar puntos anómalos sobre los cuales se puedan realizar recomendaciones médicas u optimizaciones dietarias y rutinarias. Las técnicas usadas en estos análisis incluyen análisis estadísticos de la media del nivel glucosa, la desviación estándar, coeficientes de variabilidad, el rango intercuartílico, la amplitud media de la excursión glucémica (MAGE por sus siglas en inglés) (Zheng et al., 2019, p. 3) y técnicas de machine learning como la expuestas en las secciones precedentes.



3. Materiales y Métodos

Se realizó un análisis de machine learning usando algunos de los datos obtenidos en un estudio previo publicado por Gómez et al. (2019, p. 2). Los datos corresponden a los perfiles glicémicos de un grupo de 10 pacientes durante 1 año recogidos mediante la bomba de insulina Paradigm® Veo™ que puede medir niveles de glucosa y administrar insulina y la bomba de insulina Minimed® 640G la cual además de monitorear los niveles de glucosa y administrar insulina tiene características que permiten detectar niveles bajos de glucosa e interrumpir el flujo de insulina para prevenir la hipoglicemia. Los análisis se realizaron en las fases de cargue, análisis exploratorio de datos y análisis con algoritmos de machine learning utilizando Python versión 3.10. Las librerías Pandas, Numpy, os, datetime, seaborn, matplotlib, pttiprince, researchpy, scipy y sklearn fueron utilizadas para estas tareas y la creación de las visualizaciones de los datos.

- En la fase de cargue de los datos se realizó la creación de un Dataframe de la librería pandas con las columnas Fecha, Hora, identificación anonimizada del paciente, niveles de glucosa en mg/dL, tipo de equipo utilizado. Los datos nulos fueron reemplazados usando la técnica de interpolación lineal hacia adelante por paciente, para garantizar que la tendencia de los niveles de glucosa del paciente se mantuviera en el Dataframe. Se creó la columna hora, para capturar los dos dígitos de la hora en la cual se tomó el nivel glucosa mediante la bomba de insulina y también la columna Rango en la cual niveles de glucosa menores a 70 mg/dL fueron marcados como Bajo, niveles superiores a 180 mg/dL fueron marcados como Alto y los datos entre 70-180 mg/dL fueron marcados como Normal.
- En la fase de análisis exploratorio de datos se realizó una descripción del Dataframe mediante análisis gráficos de los niveles de glucosa, rangos y horas, usando gráficos de dispersión, Rainplots (combinación de violin plot, grafico de dispersión y boxplot) y medidas de tendencia central del nivel de glucosa (conteo, media, desviación estándar, mediana, 1er cuartil, 3 cuartil y el máximo). En esta fase se realizó también una prueba de estadístico t para determinar si existían diferencias estadísticamente significativas entre los dos equipos disponibles en el Dataframe para corroborar junto con Gómez et al. (2019, p.2) que el equipo Minimed® 640G reducía los eventos de hipoglicemia; Finalmente se realizó un análisis de frecuencias de ocurrencias de los rangos determinados anteriormente comparándolos con la hora para determinar los horarios en los cuales había mayores ocurrencias de estos rangos de niveles de Glucosa.
- En la fase de análisis mediante técnicas de machine learning se realizó la clasificación de datos anómalos mediante 3 métodos: (1) Estadístico, en el cual se definió que los valores de glucosa ubicados dos desviaciones estándar por encima de la media y que al mismo tiempo fueran mayores a 180 mg/dL serían marcados como Alto, indicando que son anormalmente altos, los valores ubicados dos desviaciones estándar por debajo de la media y que al mismo tiempo fueran menores a 70 mg/dL serían marcados como Bajo indicando que son anormalmente bajos, los demás serían marcados como Normal. (2) Método de Local Outlier Factor (LOF) el cual determina por cada punto si es normal o anormal dependiendo si los puntos vecinos están cerca o lejos; para este método se usaron 20 vecinos y se definió que si un punto era marcado como anormal por el modelo de LOF y además era menor a 70 mg/dL sería marcado como Bajo, indicando que es anormalmente bajo, si el punto era marcado como anormal y además es mayor a 180 mg/dL sería marcado como Alto, indicando que es anormalmente alto y en cualquier otro caso sería marcado como Normal. Finalmente (3) se usó el método de los K



vecinos más cercanos (K-nearest neighbors KNN) para predecir anomalías, en el cual se definió una comparación de un punto con 20 vecinos, un promedio de distancia entre ellos superior a 5 para ser considerado como anómalo y (igual que en los dos métodos anteriores) se marcó como Bajo los datos considerados anómalos por el modelo que además fueran inferiores a 70 mg/dL indicando que son anormalmente bajos, se marcó como Alto los datos que fueran anómalos según el modelo y que además fueran mayores a 180 mg/dL indicando que son anormalmente Altos y los demás fueron marcados como Normal.

4. Resultados y Discusión

Con base en las consideraciones mencionadas anteriormente los siguientes resultados fueron encontrados:

Se encontró que hay picos ascendentes en los niveles de glucosa en horas del día asociadas a horas de la comida, evidenciadas en la Fig. 1.

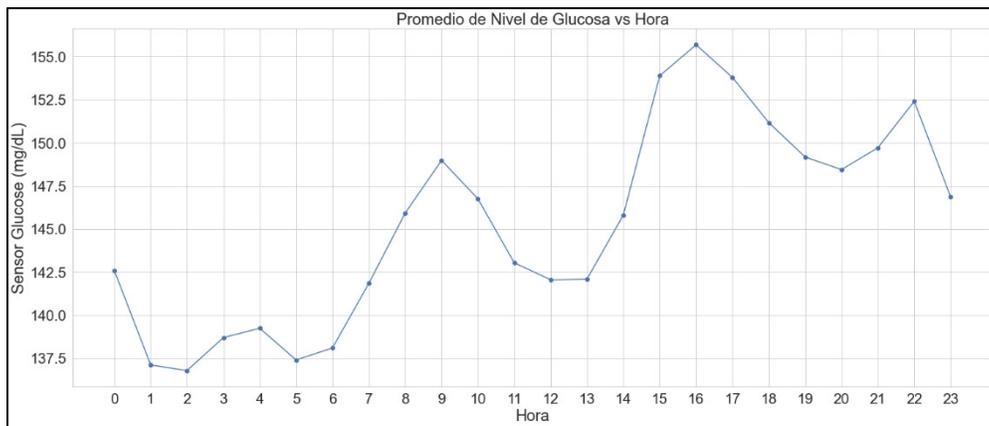


Fig. 1. Promedio de Nivel de Glucosa vs Hora

Del grafico de Rainplot por Paciente (Fig. 2) se pueden observar distribuciones sesgadas hacia la derecha (Mediana mayor a la moda) en la que la mayoría de los datos se ubican entre 100 y 200 mg/dL de glucosa. También se puede observar, mediante la amplitud del rango intercuartílico y la altura del violinplot, que hay pacientes que presentan una alta desviación estándar en sus datos.

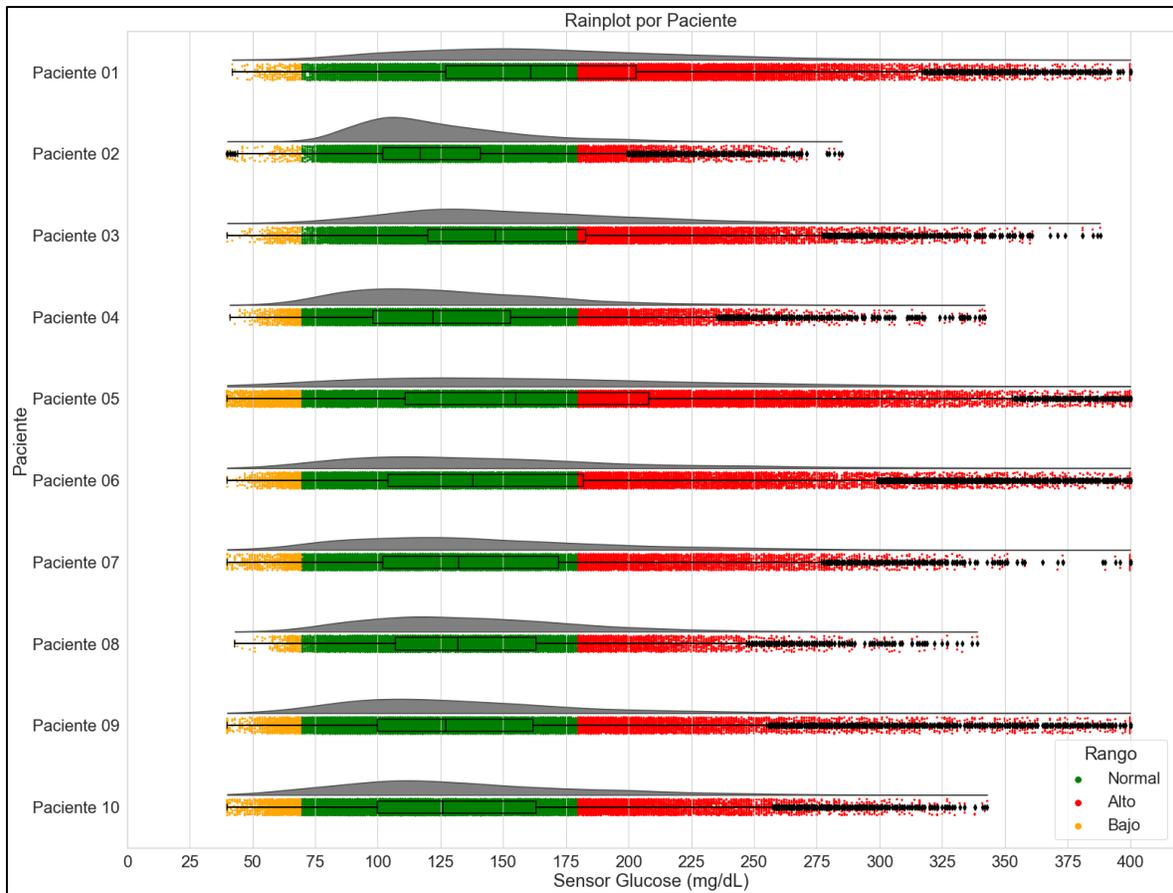


Fig. 2. Rainplot por Paciente

En la prueba de estadístico t se obtuvo un valor $p = 0.024$ usando la librería stats y un valor $p = 0.012$ usando la librería researchpy, lo que nos permite concluir que si hay diferencias estadísticamente significativas entre los sensores de glucosa.

Tabla 1. Resultados de la aplicación de modelos de Machine Learning para detección de datos anómalos

Método	Bajo	Normal	Alto	Total
Estadístico	0	232824	10807	243631
LOF	601	240820	2210	243631
KNN	3797	218460	21374	243631

En el análisis de anomalías se encontraron los resultados descritos en la Tabla 1, en la cual se puede evidenciar que el método que más anomalías logró identificar fue el método KNN, esto puede deberse a los parámetros usados en la creación del modelo. También se puede observar que el método estadístico no logró capturar ningún dato anormalmente bajo, esto debido a que el mínimo valor leído por los sensores de glucosa fue 40, pero el valor de corte usado es 34, es decir que no hay ningún dato que esté ubicado dos desviaciones estándar por debajo de la media y que al mismo tiempo sea menor a 70 mg/dL.



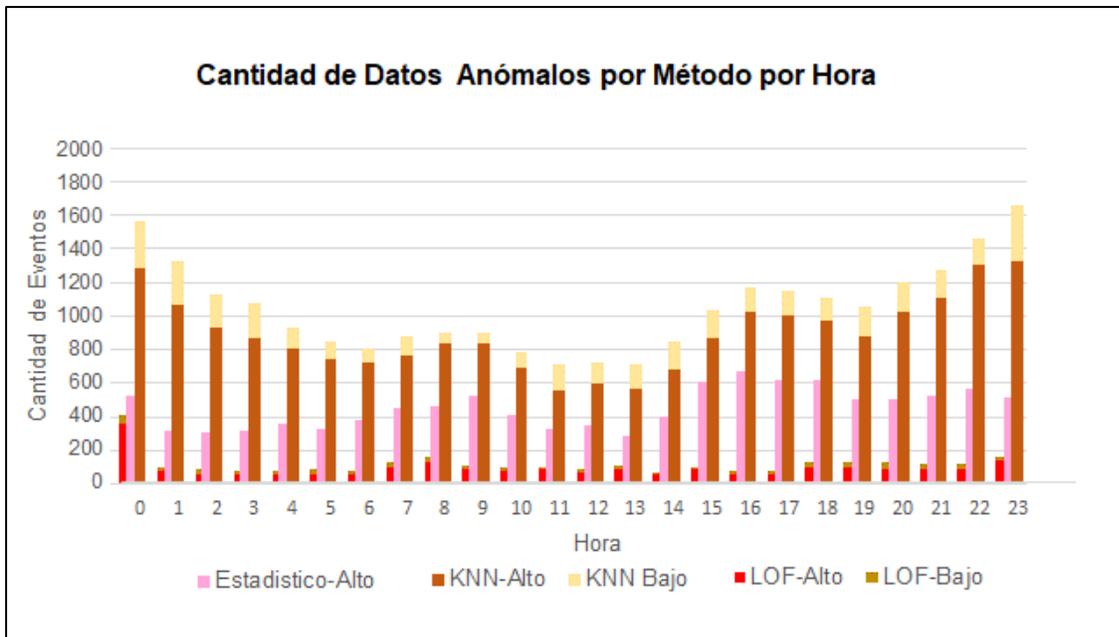


Fig. 3. Cantidad de Datos Anómalos por Método por Hora

De la Figura 3 se puede observar, además de los resultados de la Tabla 1, el hecho de que el método KNN encuentra una mayor cantidad de datos anómalos altos entre las 22 y las 0 horas, posiblemente asociados al hecho de que entre las 23 y las 2 horas es donde se presenta una mayor proporción de datos en el rango bajo (menores a 70 mg/dL), lo cual aumenta el promedio de la distancia entre los datos altos y sus 20 vecinos resultando en que la mayoría de datos por encima de 180 mg/dL en ese intervalo sean marcados como anormalmente altos por el método KNN. En el método estadístico se observa que la cantidad de datos anómalos sigue la tendencia del promedio del nivel glucosa por hora, esto se debe a que el método estadístico marca como anómalos los datos usando un valor fijo.

5. Conclusiones

- A pesar de que la mayor cantidad de datos se ubican en el rango normal de glucosa (70-180 mg/dL) las 22, 16, 17, 15 horas del día son en las cuales se presenta una mayor proporción de datos en el rango alto (mayores a 180 mg/dL); de igual manera el intervalo entre las 23 y las 2 es donde se presenta una mayor proporción de datos en el rango bajo (menores a 70 mg/dL).
- Al analizar la diferencia de medias de los resultados de la prueba de estadístico t observamos que las medias en el sensor Paradigm® Veo™ son menores a las del Minimed® 640G, con lo cual se puede concluir que hay una menor cantidad de puntos bajos en este último equipo que causa el efecto de subir la media del valor de glucosa observado con respecto al Paradigm® Veo™. Esto puede deberse a que el dispositivo Minimed® 640G tiene un algoritmo de predicción de hipoglucemias para suspender el suministro de insulina y evitar valores muy bajos de glucosa.



- El método KNN logra capturar una mayor cantidad de anomalías en el Dataframe adaptándose a los valores de glucosa del intervalo y prediciendo con base en las mediciones anteriores y posteriores si un dato es anormalmente alto o bajo.

6. Referencias

- Bando, H.; Ebe, K., Kato, Y.; Bando, M.; Yonei, Y. (2018). Investigation of blood glucose profile by continuous glucose monitoring (CGM). *Endocrinology Research and Metabolism*. Vol. 2 Mo. 1:10. 1-4.
- Blázquez-García, A.; Conde, A.; Mori, U.; Lozano, J. A. (2021). A Review on Outlier/Anomaly Detection in Time Series Data. *ACM Computing Survey*, (54), No. 3. 1 - 33. DOI: <https://doi.org/10.1145/3444690>
- Castillo J. M. et al. (2019). Impacto del alto costo relacionado con la diabetes Mellitus en el sistema de salud en Colombia. *REDIIS / Revista De Investigación E Innovación En Salud*, 2, 82-91.
- Chandola, V.; Banerjee, A.; Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Survey*, (49), No 3. 1 – 58. DOI: 10.1145/1541880.1541882
- Collins English Dictionary. (s.f. a) Definition of Univariate. Recuperado de <https://www.collinsdictionary.com/dictionary/english/univariate>
- Collins English Dictionary. (s.f. b) Definition of Multivariate. Recuperado de <https://www.collinsdictionary.com/dictionary/english/multivariate>
- Cortés, C.; Vapnik, V. (1995). Support-Vector Networks. DOI: 10.1007/BF00994018
- Gomez, A. M, et al. (2019). Impact of sensor-augmented pump therapy with predictive low-glucose management on hypoglycemia and glycemic control in patients with type 1 diabetes mellitus: 1-year follow-up. DOI: <https://doi.org/10.1016/j.dsx.2019.07.024>
- He, Z.; Xu, X.; Deng, S. (2002). Discovering cluster-based local outliers. *Elsevier, Pattern Recognition Letter*, (24). 1641 - 1650. DOI: 10.1016/s0167-8655(03)00003-5
- Hilal, W.; Gadsden, S. A.; Yawney, J. (2021). Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances. *Elsevier*. 1 – 34. DOI: <https://doi.org/10.1016/j.eswa.2021.116429>
- IDF Diabetes Atlas 10th edition (2021) Consultado el 1 de junio de 2022 en https://diabetes-atlas.org/idfawp/resource-files/2021/07/IDF_Atlas_10th_Edition_2021.pdf
- Jain, P.; Joshi, A.; Mohanty, S. P. (2021). Everything you wanted to know about continuous glucose monitoring. 1-51.
- Jiang, M. F.; Tseng, S. S.; Su, C. M. (1999). Two-phase clustering process for outliers detection. *Elsevier, Pattern Recognition Letter*, (22). 691 - 700. DOI: 10.1016/s0167-8655(00)00131-8
- Jiang, Y. S.; Song, X.; Wang, H.; Han, J. J.; Li, Q. H. (2006). A clustering-based method for unsupervised intrusion detections. *Elsevier, Pattern Recognition Letter*, (27). 802- 810. DOI: 10.1016/j.patrec.2005.11.007
- Ng, A. y Russel, S. (2000). Algorithms for Inverse Reinforcement Learning. *ICML*. 663-670. DOI: <https://dl.acm.org/doi/10.5555/645529.657801>
- Pang, G.; Shen, C.; Cao, L.; van den Hengel, A. (2021). Deep Learning for Anomaly Detection: A Review. *ACM Computing Survey*, (54), No. 3. 1 - 38. DOI: <https://doi.org/10.1145/3439950>
- Pang, G.; Ting, K. M.; Albrecht, D. (2015). LeSiNN: Detecting anomalies by identifying least similar nearest neighbours. *2015 IEEE 15th International conference on Data Mining Workshops*. 623 – 630. DOI: 10.1109/icdmw.2015.62
- Ramaswamy, S.; Rastogi, R.; Shim, K. (2000). Efficient Algorithms for Mining Outliers from Large Data Set. 427 – 438. DOI: 10.1145/342009.335437
- Saha S, Al-Rifai RH, Saha S. Diabetes prevalence and mortality in COVID-19 patients: a systematic review, meta-analysis, and meta-regression. *J Diabetes Metab Disord*. 2021 Mar 31;20(1):939-950. DOI: 10.1007/s40200-021-00779-2.



- Sugiyama, M; Borgwardt, K. M. (2013). Rapid distance-based outlier detection via sampling. Recuperado en: <https://papers.nips.cc/paper/2013/hash/d296c101daa88a51f6ca8cfc1ac79b50-Abstract.html>
- Zhang, K.; Hutter, M.; Jin, H. (2009). A new local distance-based outlier detection approach for scattered real-world data. 813 – 822. DOI: 10.1007/978-3-642-01307-2_84
- Zheng, F.; Jalbert, M.; Forbes, F.; Bonnet, S.; Wojtuszczyzn A.; Lablanche, S.; Benhamou, P. Y. (2019). Characterization of daily glycemic variability in subjects with Type 1 Diabetes using a mixture of metrics. Diabetes Technology and Therapeutics. 1-40 DOI: 10.1089/dia.2019.0250

Sobre los autores

- **Anderson Mosquera Ruiz:** Ingeniero de Petróleos, Estudiante de Ingeniería de Sistemas. amosque34443@universidadean.edu.co
- **Maira Alejandra García Jaramillo:** Ingeniera de Sistemas, Máster en Tecnologías de la Información, Doctora en Tecnología de la Universitat de Girona. Profesora titular la Universidad EAN. Directora del grupo de investigación y desarrollo tecnológico ONTARE. magarcia@universidadean.edu.co

Los puntos de vista expresados en este artículo no reflejan necesariamente la opinión de la Asociación Colombiana de Facultades de Ingeniería.

Copyright © 2022 Asociación Colombiana de Facultades de Ingeniería (ACOFI)

