Sistema de perfilamiento para mejora de productividad usando machine learning

Nicolas Ricardo Enciso, Daniel Esteban Rodríguez, Camilo Andrés Pinilla

[1]

I. Resumen

Con las dinámicas laborales actuales, tener un estilo y ritmo de trabajo funcional se hace una tarea cada vez más importante. Se propone un sistema que ayude al usuario a obtener las horas de sueño y de trabajo recomendados de acuerdo a un análisis en donde se usa machine learning supervisado con métodos de clustering y k vecinos entrenados con datos reales de personas determinando las horas de sueño y descanso , además de otras variables que mejor se han ajustado a su productividad, resultando en un perfilamiento del usuario en un grupo de datos de usuarios previos, categorizando a partir de su estado de ánimo, gustos y afinidades las horas de descanso y trabajo recomendadas en un día. Se obtienen horas acordes con el perfil del usuario, así como una precisión aceptable en el perfilamiento del usuario.

***Palabras Clave*—Clustering, estado anímico, k vecinos perfilamiento.**

II. Introducción

El establecimiento de un estilo de horas de trabajo y descanso es una tarea de prueba y error, donde se estiman y destinan los tiempos de acuerdo con la urgencia del asunto, así como a la predisposición de la persona a la realización de las tareas que debe cumplir. Es por esta razón que se pretende implementar una herramienta que ayude a las personas a estimar los tiempos de descanso y de trabajo que puede destinar al día, de acuerdo a su estado de ánimo, la edad, el género y demás variables que influyen de forma directa e indirecta en la predisposición de la persona frente a la realización de las tareas pendientes, así como a la eficiencia de su trabajo, con la ayuda de datos dados por personas reales, quienes brindaron información acerca de su perfil y las horas de descanso y trabajo que destinan diariamente y que han resultado de forma positiva en los resultados de sus tareas. La idea principal se basa en la forma de pomodoro, donde se estiman unos tiempos de descanso frente a unos tiempos de trabajo, teniendo en mente que una persona necesita tener un equilibrio entre éstas dos tareas, para que pueda tener una buena eficacia en la realización de sus deberes, así como una estabilidad emocional.

El sistema propone el aprovechamiento de los datos dados por personas que ya cuentan con un estilo de trabajo determinado, de manera que el usuario del sistema pueda ser perfilado de acuerdo a una agrupación por medio de métodos de clustering de personas según sus características, que ayudan a determinar qué estilo de persona se ajusta más a la que el usuario presenta, para posteriormente hacer un análisis de los puntos que pertenecen al grupo al que fue clasificado el usuario, con ayuda del método de k vecinos y múltiples comparaciones de clusters de a dos variables, precisando las horas de descanso y de trabajo que presentan tendencia en el grupo y vecinos en el que el usuario ha sido perfilado, así como poder determinar un tipo de posible perfil que es cercano al del usuario.

III. Planteamiento del problema

A. Contextualización del problema

En el diario vivir de la población estudiantil y trabajadora, encontramos que esta desempeña múltiples y variadas actividades de las cuales surge la problemática sobre el rendimiento general que tiene una persona al distribuir incorrectamente su tiempo en descanso y trabajo, haciendo que la improductividad y la falta de disposición por parte del usuario afecte de manera directa la ejecución de los deberes, y que la tarea de encontrar la asignación de horas sea una constante de prueba y error donde no se pueden determinar con certeza los estimados necesarios de horas para poder cumplir con los deberes sin afectar el bienestar de la persona. Dado a la relación existente entre el estado anímico, físico, edad, género entre otras variables con la productividad de una persona, es posible determinar un estimado de lo que podría servir como un estilo de productividad al usuario de acuerdo con una similitud de características con otras personas que se parezcan en sus variables y ya tengan un estilo fijo de trabajo, siendo ese estilo posiblemente bien acomodado en el del usuario.

B. Definición del problema

Dado lo anterior encontramos la problemática de encontrar un estilo de trabajo, más concretamente en el estimado de horas diarias necesarias en dormir y trabajar que funcione en la medida de que la persona pueda responder a sus obligaciones sin afectar de manera colateral su bienestar, como lo puede ser el estado de ánimo y físico. Adicionalmente, se tiene que las personas necesitan de constantes pruebas donde intentan usar determinados horarios de descanso y trabajo según vayan dando resultados óptimos, haciendo que se prolongue el tiempo de desequilibrio entre el cumplimiento de los deberes, con el bienestar de la persona.

[1]Nicolas Ricardo Enciso: nricardoe@unal.edu.co, estudiante de Ingeniería de Sistemas y Computación, Universidad Nacional de Colombia.

Daniel Esteban Rodríguez: daerodriguezme@unal.edu.co, estudiante de Ingeniería de Sistemas y Computación, Universidad Nacional de Colombia.

Camilo Andres Pinilla: capinillab@unal.edu.co, estudiante de Ingeniería de Sistemas y Computación, Universidad Nacional de Colombia.

Se pretende asignar un tiempo en horas de sueño o descanso y de trabajo que mejor se acomoden a la persona de acuerdo a la similitud de sus características con la de otros usuarios que ya han determinado sus horarios de trabajo y descanso, de manera que la persona obtenga un sugerido más preciso de acuerdo a su perfil, que pueda resultar en un inicio de mejora de productividad.

IV. ANTECEDENTES

En marketing y publicidad, es común clasificar el perfil de los clientes de una empresa en varios grupos, de manera que se tengan catalogados determinadas preferencias que los clientes tengan respecto a ciertos productos. En años recientes se ha estado implementando machine learning en la clasificación de grupos de clientes,[1] con la idea de poder sectorizar sus gustos, para poderles ofrecer ya sea productos especiales que se acomodan a ese perfil, servicios, planes etc. Así mismo, la clasificación de perfiles de clientes, permite identificar nichos de mercado no explorados por nuevos perfiles de personas que no se tenían en planes. Se presenta entonces una relación con el sistema propuesto en el presente artículo, en la medida en que la funcionalidad de sugerencia de tiempos de trabajo para los usuarios, usa las ideas del perfilamiento y la clasificación de las personas, de acuerdo a unos parámetros no numéricos, con el ánimo de sugerirles información de acuerdo a dónde se ubiquen en la clasificación del perfilamiento hecho a partir de datos anteriores que ayudan a entrenar previamente la clasificación del perfilamiento.

De igual manera se ha implementado el perfilamiento en el reconocimiento y etiquetado de personalidades de acuerdo a un análisis de perfiles en redes sociales, donde por ejemplo, se implementa un método llamado LDL (labeled distribution learning), un estudio demuestra que basados en las 5 principales personalidades, se es posible clasificar a una persona en un tipo de perfil de personalidad, relacionando el estado emocional, anímico y de preferencias con el tipo de personalidad que tiene una persona [2]. En resumen, con un análisis de la cuenta personal de alguien en una red social, es posible a través de machine learning asignarle una personalidad, con el ánimo de agruparla en una categoría para poder tener mejor precisión al momento de sugerirle productos, mostrarle publicidad. Mostrándose así la relación directa entre el estado de ánimo de una persona con su personalidad, el enfoque que se da al presente artículo, respecto a usar como variables de entrada unos rasgos de personalidad traducidos a intereses que tenga el usuario, los cuales inciden en los estados anímicos de la persona.

Por otra parte, se ha podido establecer, con un estudio, que la productividad de una persona puede ser mejorada con el uso de técnicas de trabajo basadas en pomodoro [3]. La idea principal se desarrolla en la minimización de las distracciones de los desarrolladores de software que trabajan bajo las metodologías ágiles, teniendo en cuenta que en el desarrollo de software, los tiempos de entrega de producto son críticos, y las eventualidades son ineludibles. Se propone entonces aplicar de forma personalizada la técnica pomodoro en los horarios de trabajo de un desarrollador que trabaja bajo la metodología ágil. El estudio se centra en el tratamiento de las distracciones que ocurren en los tiempos de trabajo, basado en el tipo de tareas que debe desarrollar, construyendo sus propias reglas categorizando los tipos de interrupciones que afectan su productividad. Se obtienen así en el estudio, incrementos en la productividad debido directamente al decrecimiento de las distracciones, mostrando así que las técnicas basadas en pomodoro para la mejora de los tiempos productivos funcionan de forma importante, por lo cual en el presente proyecto, la implementación de algunas ideas basadas en la técnica de pomodoro pueden ayudar de forma significativa en mejorar los tiempos productivos que se le pueden sugerir a una persona.

Otro campo relacionado a la clasificación de personas, es decir al perfilamiento, hace referencia a la posibilidad de relacionar cuentas de personas en redes sociales, con un perfil de edades característico, capaz de detectar a través del análisis de sentimientos, en qué rango de edad está esa persona. De ésta manera se establece según el estudio, que se tiene una correlación entre los gustos, intereses de una persona, con la edad de la misma [4]. Adicionalmente con un análisis de los comentarios que una persona hace acerca de diversos temas en redes sociales, se es capaz de establecer de forma certera la edad aproximada de quien expresa el comentario, brindando así una herramienta capaz de sugerir de forma precisa publicidad.

Es así como se tiene entonces relación entre los intereses de una persona, ya sea a partir de publicaciones o datos dados por la misma persona, con su edad aproximada, con un análisis de machine learning, mostrando así que la variable de edad puede ser determinante a la hora de perfilar los intereses de una persona.

Un área transversal, relacionado es el de data mining, el cual es estrechamente relacionado al campo del machine learning. Un estudio demuestra que con un análisis detallado de las compras de una persona en internet, se puede mejorar el nivel de compras debido a un incremento en la satisfacción del cliente específicamente en el comercio online [5]. Lo anterior muestra que con una retroalimentación de un previo análisis de grandes cantidades de datos, se puede mejorar la experiencia en compras con patrones encontrados en los datos de una persona. Se ha establecido que los algoritmos que usan machine learning para sugerir productos a los clientes, basados en datos de sus anteriores transacciones, resultan ser muy efectivos, concretamente en el caso de Netflix, y su sistema de sugerencia de series y películas [6]. El sistema toma la información del usuario respecto a los géneros que ha visto, los horarios, y la información personal del usuario, para clasificarlo con un perfil determinado de usuario, un etiquetado, donde se le van sugiriendo productos similares a los que se han estado

consumiendo, basados en la clasificación del perfilamiento de sus gustos, con el añadido de tener una constante retroalimentación por parte del usuario hacia el algoritmo de machine learning, debido a que se va revisando si de las sugerencias que se le han dado, cuales descarta y cuales toma, permitiendo que se vaya afinando en cada nueva instancia, el perfil de usuario y sus gustos respecto al material audiovisual que consume en la plataforma.

Los algoritmos de sugerencias también son aplicables a los análisis de datos con redes sociales como se muestra en un estudio[7], donde por medio de clustering a partir de k vecinos, es posible delimitar de manera precisa, haciendo correlaciones con perfiles de información de personas similares, mejorando de forma considerable los anteriores métodos de recomendaciones que se basaban en el historial de búsquedas y demás información anterior de una persona [8].

Aporte de enfoque del proyecto

Como se muestra en la contextualización y trabajos previos relacionados con el estudio de estados de ánimo, perfilamiento de personas, y estudio de técnica pomodoro, el área de aplicación de machine learning desde el clustering, para la clasificación de personas tiene gran precisión y múltiples funcionalidades. El presente artículo, pretende aportar en la medida de aplicar en un nuevo campo las ya mencionadas técnicas, en la estructuración de tiempos acordes a variables personales y subjetivas como lo es el estado del ánimo y los intereses personales, adicional de otras variables como las horas de descanso, la edad donde previamente se mostró con los artículos citados la posibilidad de conexión entre la edad de una persona y sus intereses, dando pie a desarrollar a través de la clasificación por clustering de personas, un sistema de sugerencia de tiempos de trabajo y de sueño, ya que, los intereses y el estado de ánimo como primeras variables, establecen una buena primera instancia de sugerencia de tiempo de trabajo, aprovechando las bondades de la aplicación de machine learning, como lo son la adaptación y la escalabilidad a nuevas variables y métodos que ayuden a precisar aún más las sugerencias que se le den al usuario del sistema, aprovechando en gran medida el uso de grande datos de personas reales que han encontrado su forma perfecta de estimación de horas de trabajo y descanso, de manera que se amplía la precisión del presente análisis que se plantea, teniendo así un sistema escalable que brinda una nueva alternativa de estudio de perfilamiento de personas con miras a sugerir tiempos de descanso y trabajo acordes.

V. METODOLOGÍA

A. *Descripción detallada del enfoque propuesto*

En una primera instancia, se propone la recolección de una gran cantidad de datos de personas que tiene ya fijado un estimado de horas de sueño o descanso frente a unas horas de trabajo, además de otras variables relacionadas con características personales, de manera que se tenga un amplio rango de variables que permitan precisar aún los grupos que se pueden formar con los clusters.

Teniendo así los datos de personas, se procede a una sectorización de esos datos, a modo de entrenamiento del sistema, de manera que se puedan categorizar grupos de personas que cuentan con unas determinadas características similares, haciendo así una asociación de parámetros que tienen en común un grupo de personas.

Se sigue una idea en la cual las personas que tienen características personales similares tienen por consiguiente un estilo de horas de trabajo y de descanso similares, por lo que personas que sean clasificadas en el mismo grupo, se les podrá sugerir los tiempos que ése grupo de personas vienen usando. El enfoque es entonces, el uso de una clasificación de grupos de personas con características comunes, es decir un perfilamiento, en el cual la pertenencia o no a un grupo característico, comparte los tiempos de trabajo y descanso que mejor se pueden acomodar a quienes lleguen a pertenecer a dicho grupo, teniendo así un inicial estimado de asignación de tiempos mucho más preciso que ayudan en gran medida a los nuevos usuarios a obtener un tiempo acorde a sus características.

B. *Datos de entrenamiento*

Los datos de entrenamiento son obtenidos por medio de un cuestionario de preguntas entre las que estaban de opción múltiple, única selección, y respuesta abierta numérica, el cual es compartido por redes sociales donde mayoritariamente estudiantes universitarios responden, obteniendo así un estimado de 224 perfiles de personas.

El cuestionario se basaba en 8 preguntas las cuales dirigían a la obtención de las variables usadas en la ubicación de las personas. Las variables evaluadas son:

a. Edad
b. Horas de sueño/descanso promedio
c. Estado de ánimo mayoritario
d. Gustos o intereses personales
e. Horas productivas diarias
f. Género
g. Ocupación
h. Estado físico mayoritario

Las variables a, b, e fueron numéricas enteras abiertas, las variables c, f, g y h fueros de selección múltiple, de manera que luego fueron discretizadas según la cantidad de opciones disponibles en cada variable, como se muestra en la TABLA 1 con el caso de la variable estado de ánimo y su discretización:

TABLA 1. Discretización de variable estado de ánimo

Muy feliz	1
Feliz	0.8
Normal	0.6
Triste	0.4
Muy triste	0.2

En el caso de la variable gustos, al ser de múltiple selección se procede a hacer una discretización binaria, es decir, para cada opción de la variable, se ponía un valor de 1 si era seleccionada la opción y 0 si no era seleccionada, teniendo así un vector de respuestas como variable.

Los datos son almacenados en formato csv, obteniendo un total de 224 datos recolectados de personas reales, usando éste como el insumo de información para el entrenamiento del sistema.

B. Análisis de datos y entrenamiento de machine learning

Teniendo listo el insumo de entrada de datos de entrenamiento, se continua con la construcción de los algoritmos de machine learning a usar: clustering DBSCAN, k vecinos. En una primera instancia se hace la construcción de dos modelos de clustering: en el primero se hace una implementación propia usando como variable principal de clasificación la variable gustos. La implementación y todo el sistema es escrito en Python.

C. Descripción método de Clustering

Dadas las condiciones relativas al tipo de datos al que se quiere hacer clustering, se propone el método *DBSCAN*, el cual presenta los siguientes atributos:

- *Parámetros:* Tamaño de la vecindad
- *Escalabilidad:* Muy alta, n muestras, asociación media por n clusters
- *Caso de uso:* Tamaño de clusters no parejo
- *Métricas:* Distancias entre los puntos más cercanos

Este método define a los clusters como regiones densas de datos que están separadas por otras regiones que presentan baja densidad. Un cluster está definido como el conjunto maximal de puntos conectados, las formas de los clusters pueden ser de tipo arbitrario [9].

Definición de densidad:

$\varepsilon - Neighbourhood$ - Objetos que se encuentran dentro de un radio epsilon de un objeto específico

"High density" $\varepsilon - \boldsymbol{Neighbourhood}$ de un objeto que contenga al menos un número mínimo $MinPts$ de objetos.

(1)

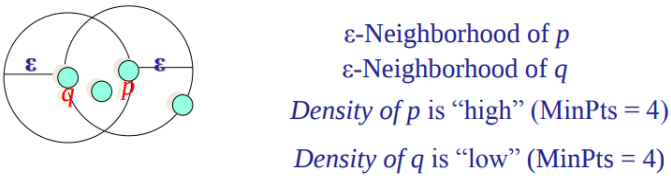

Fig 1. Descripción epsilone vecinos en densidad.

Core, border & outlier.

- *Core point:* Si tiene más de la cantidad especificada por MinPts dentro del rango Eps, estos puntos pertenecen al interior de un cluster.
- *Border point:* Si tiene menos puntos que los especificados en MinPts dentro de la distancia Eps, pero está en la vecindad de un *core point*
- *Noise point:* Es cualquier punto que no sea un *core point* o un *border point*

(2)

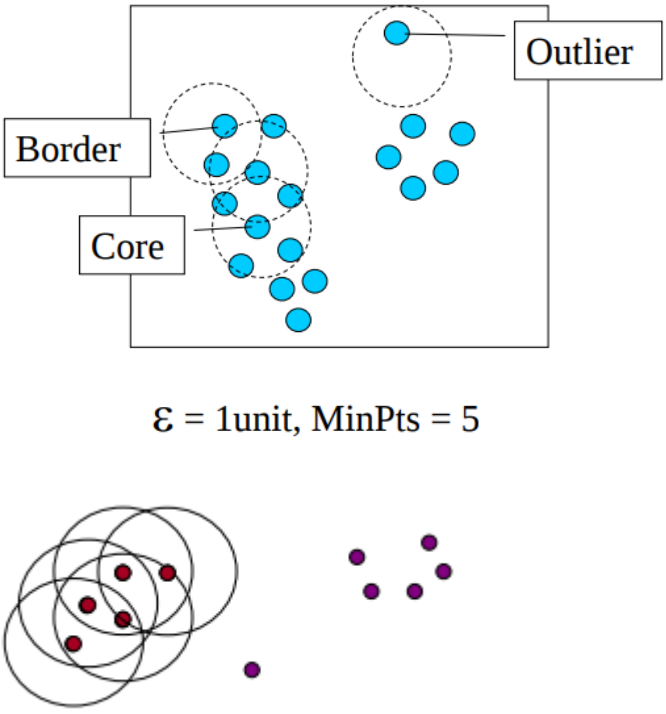

Fig 2. Descripción gráfica del funcionamiento de clustering con core point, border y outlier.

Con lo anterior descrito, el algoritmo se desarrolla de la siguiente manera [9] :

(3)

```
for each o ∈ D do
    if o is not yet classified then
        if o is a core-object then
            collect all objects density-reachable from o
            and assign them to a new cluster.
        else
            assign o to NOISE
```

Fig 3. Descripción algoritmo de clustering con core, border y outlier.

Los clusters entonces son clasificados de acuerdo con revisiones ajustando los parámetros de épsilon y cantidad de puntos en clusters, obteniendo así una clasificación de agrupamiento de personas con elementos característicos similares, teniendo lista la categorización de cualquier dato entrante, ayudando a precisar aún más las respuestas de horas de descanso y trabajo con el uso de otro método de machine learning como lo es k vecinos.

Los clusters que usan la librería DBSCAN de scikitlearn son promediados sus resultados con los de implementación propia que usa la variable gustos, para luego ser aún más precisada con el uso del método de k vecinos, haciendo uso de todas las variables disponibles, aumentando el rango de clasificación correcta del nuevo dato entrante a ser clasificado.

El nuevo dato entrante es entonces ubicado en uno de los grupos de los clusters y clasificado igualmente por k vecinos, en el presente caso se tienen 7 puntos vecinos, promediando los resultados de horas de sueño y horas de trabajo, para finalmente dar al usuario la cantidad de horas de sueño y horas de trabajo que debería empezar a implementar teniendo en cuenta que personas con perfiles similares usan ya ese rango de horas de forma exitosa.

Estructura general del desarrollo de la solución del problema:

1. Recolección de información de personas
2. Discretización de variables de data set construido
3. Entrenamiento de algoritmos de clustering con implementación propia usando como variable principal gustos, y construcción de clusters a partir de todas las posibles combinaciones de a dos variables disponibles en la data set usando librería de scikit learn.
4. Entrenamiento de algoritmo de implementación propia de k vecinos haciendo uso de todas las variables, teniendo como métrica la distancia euclidiana.
5. Clasificación final de grupos de datos combinando todos los algoritmos.
6. Entrada de nuevo dato y clasificación del mismo en un grupo de los clusters y k vecinos.
7. Determinación de las horas promedio de descanso y trabajo de todos los algoritmos respecto al grupo y vecinos en los que es acomodado el nuevo dato
8. Salida de las horas de sueño y horas de trabajo que el nuevo dato entrante debería usar.

D. Proceso de experimentación

La calibración de los resultados, la clasificación de perfilamiento de los datos de entrenamiento, así como las métricas usadas, se basaron en constantes revisiones, variando los parámetros principales de cada uno de los algoritmos de clasificación implementados: clustering implementación propia, clustering de dos dimensiones y k vecinos.

Para el caso de los clustering de tipo DBSCAN, así como el de implementación propia, la experimentación se basó en la variación de la épsilon, la distancia de radio de clasificación de puntos cercanos por densidad, así como de minPts correspondiente a la cantidad mínima de puntos dentro del radio de clasificación dado por la épsilon. La búsqueda del ideal se basó en obtener un mínimo de 2 clusters y un máximo de 9, ya que para la cantidad de datos de entrenamiento y un análisis previo de la tendencia de selección de cada variable sugería no tener demasiados grupos de clasificación con el ánimo de minimizar la cantidad de posibles puntos catalogados como ruido, y, en el caso de un mínimo de 2 clusters se determinó con la idea de poder tener una variedad mínima de grupos con el ánimo de dar sentido a hacer la clasificación de los puntos y que pudieran ser catalogados en un grupo catalogado con características lo más similares posibles.

Para el caso de la cantidad mínima de puntos, se fue ajustando de acuerdo con los resultados de cantidad de clusters que se iban determinando, es decir, la cantidad de puntos mínimos estaba ligada a la cantidad de clusters que iba apareciendo.

En el caso de k vecinos, se usó como métrica la distancia euclidiana, y, posteriormente, k como 7, debido a que fue la que mejor se acomodaba a una cantidad similar de clusters en promedio que iban dando los clusters.

VI. RESULTADOS OBTENIDOS

A. Análisis de resultados

Se obtuvo una clasificación acertada en las sugerencias de horas de sueño y horas de trabajo, de acuerdo con el perfilamiento que el sistema lanza como salida. Si bien los resultados no pueden ser medibles de acuerdo con una métrica de calidad, debido a la subjetividad de la precisión de los datos dados como resultados, si pueden ser revisados como de acuerdo al perfil ingresado en pruebas, debido a que las horas de trabajo y de sueño obtenidas son consistentes con lo que el usuario ha ido usando en su estilo de trabajo, con lo que se tiene una correlación importante que puede dar como conclusión la buena precisión del sistema respecto a las sugerencias de datos dados.

En cuanto al perfilamiento de acuerdo con las demás variables, se tiene cierta correlación especialmente en la edad del usuario. En cuanto al género se tiene en algunos casos disparidades con los vecinos, que, si bien no es algo que afecte de manera directa el resultado, si muestra la necesidad de usar una mayor cantidad de datos que, a través de una densidad de datos mayor, podría dar con una mejor precisión del género.

B. Discusión de resultados

La subjetividad de los resultados dificulta hacer un preciso estimado de la exactitud obtenida, ya que el criterio de efectividad es dado por el mismo usuario, quien es quien decide si es funcional o no el resultado de las horas de trabajo y las horas de descanso que el sistema da como resultado. También se nota la necesidad de obtener más datos que ayuden a sacar aún más provecho del manejo de la densidad de puntos que posee la clasificación por medio de clustering, por lo que algunos datos del perfilamiento del usuario mantienen cierta distancia respecto a los esperados.

VII. RECOMENDACIONES DE MEJORAMIENTO

Con los experimentos ejecutados, se puede observar la necesidad de tener una mayor cantidad de datos, de manera que esto puede aumentar de manera considerable la efectividad en cuanto a la clasificación de los puntos ya se por medio del algoritmo de clustering, o por medio de k vecinos. Esto se puede notar en los casos de clustering de a dos variables en donde la cantidad de puntos catalogados como ruido eran elevados, dando a entender una dispersión considerable de los puntos de entrenamiento a tratar.

Por otro lado, al usar varias implementaciones de algoritmos de clasificación y distintos tipos de análisis con diferentes configuraciones de variables, se obtuvieron mejoras en la consistencia de los datos arrojados como resultado, por lo que se abre la posibilidad de mejoramiento a través de la combinación de varios métodos diferentes y posteriormente, una centralización de los resultados, haciendo que la precisión de los resultados sea aún mayor, y que se ayuden a minimizar casos de dispersión alta entre los resultados esperados y los resultados obtenidos en el sistema.

VIII. CONCLUSIONES

La clasificación o perfilamiento de personas con entrada un grupo de variables, hacen posible ubicar a determinados individuos dentro de un grupo ya establecido, etiquetándolos con determinadas características propias del grupo en el que se ubica. Se muestra que por medio de la clasificación por clustering y por k vecinos es posible ir clasificando personas no sólo para obtener las horas de sueño y trabajo que mejor se acomodan, sino a la posibilidad e poder referenciar a una persona de acuerdo a otras características que ayuden de determinar la pertenencia o no de un individuo a un grupo de personas, lo que abre la posibilidad de poder implementar las mismas ideas desarrolladas en el presente artículo, en otros campos que requieran perfilar personas de acuerdo a los datos de personas que ya han sido registradas, pudiendo así determinar qué patrones de características y comportamientos pude presentar la persona sin tener aún confirmación de la persona.

Se concluye también que la combinación de varios métodos de clasificación de machine learning y varias formas de implementación, pueden ayudar a mejorar la precisión de los datos en cuanto a la consistencia de los mismos respecto a lo esperado.

Por último, se observa la necesidad de tener un gran data set capaz de presentar de forma óptima las densidades que sacan aún más provecho de los algoritmos de clasificación basados en clustering, evitando así una continua verificación y cambios en la configuración de los parámetros propios de los algoritmos como lo son el épsilon de distancia entre puntos, y la cantidad de puntos mínimos incluidos en el rango de medición determinado para épsilon. Lo anteriormente mencionado también es aplicable a la clasificación por k vecinos, debido a que, si se tienen concentraciones de puntos de manera evidente, se facilita la clasificación y categorización de nuevos datos entrantes con el fin de poder dar resultados basados en las características propias del grupo en el cual es ubicado.

IX. REFERENCIAS

[1] Indrail Bose, Radha K. Mahapatra, "Business data mining a machine learning perspective", en *Information & management 39 (2001) 211-225*, 2001.

[2] Di Xue , Zheng Hong , Shize Guo , Liang Gao , Lifa Wui , Jinghua Zheng , y Nan Zhao, "*Personality Recognition on Social Media With Label Distribution Learning",* en *IEEE Access Volumen 5* , 2017.

[3] Mintra Ruensuk, Stamford University, Bangkok Tailandia, "*An implementation to reduce internal/external interruptions in Agile Software Development Using Pomodoro Technique ",* en *Computer and Information Science (ICIS), 2016 IEEE/ACIS 15th International Conference,* 2016.

[4] RITA GEORGINA GUIMARÃES, RENATA L. ROSA, DENISE DE GAETANO, DEMÓSTENES Z. RODRÍGUEZ, (Senior Member, IEEE), AND GRAÇA BRESSAN, "*Age Groups Classification in Social Network Using Deep Learning",* en Minas Gerais Universidad de Sao Paulo, Universidad Federal de Lavras, FAPEMIG Minas Gerais agencia estatal de investigación *IEEE Access Volumen 5*, 2017.

[5] Ronald E. Goldsmith , "*Explaining and Predicting Consumer Intention to Purchase Over the Internet: An Exploratory Study",* en *Journal of marketing theory and practice* , 2002.

[6] Giovanny M. Tarazona B, Juan S. Chávez L, Roberto Ferro E, "*Modelación de sistemas de recomendación aplicando redes neuronales artificiales"* en *Visión electrónica Universidad Distrital Francisco José de Caldas*, 2013.

[7] Taiping Lai, Xianghan Zheng, "*Machine Learning Based Social Media Recommendation",* en *Spatial Data Mining and Geographical Knowledge Services (ICSDM), 2015 2nd IEEE International Conference*, 2015.

[8] Q. Gao, L. Xin, "Products Recommend Algorithm Based on Customer Preference Model and Affective Computing," in Proc. CCC'2010, 2010: 2891-2896.

[9] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, *"A density-Based Algortihm for Discovering Clusters"* , en *KDD proceding, AAAI, Universidad de Munich*, 1996.