



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería Informática



**TFG del Grado en Ingeniería
Informática**

SmartBeds

**Aplicación de técnicas de
minería de datos para la
detección de crisis epilépticas**



Presentado por José Luis Garrido Labrador
en Universidad de Burgos — 20 de mayo
de 2019

Tutores: Álgvar Arnáiz González y José
Francisco Díez Pastor



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería Informática



Dr. D. Álgvar Arnáiz González, profesor del departamento de Ingeniería Civil, área de Lenguajes y Sistemas Informáticos.

Expone:

Que el alumno D. José Luis Garrido Labrador, con DNI **TODO**, ha realizado el Trabajo final de Grado en Ingeniería Informática titulado **Smart-Beds**

Aplicación de técnicas de minería de datos para la detección de crisis epilépticas.

Y que dicho trabajo ha sido realizado por el alumno bajo la dirección del que suscribe, en virtud de lo cual se autoriza su presentación y defensa.

En Burgos, 20 de mayo de 2019

Vº. Bº. del Tutor:

Vº. Bº. del co-tutor:

Dr. D. Álgvar Arnáiz González

Dr. D. José Francisco Díez Pastor

Resumen

La epilepsia es una enfermedad que provoca crisis repentinas con convulsiones violentas y pérdidas del conocimiento. Ante estas situaciones es necesario suministrar unos primeros auxilios, sin embargo, durante las sesiones nocturnas es difícil detectar y tratar a tiempo estas crisis que pueden provocar daños graves a los pacientes que sufren esta enfermedad.

En este trabajo se investiga, a partir de datos reales, métodos que puedan detectar situaciones de crisis en tiempo real de tal manera que se puedan suministrar las atenciones necesarias con la mayor celeridad posible.

Además de esto se ha creado una API REST y una página web con una gestión en tiempo real de camas con el estado actual del paciente y las probabilidades de una situación de crisis.

Descriptores

Crisis epilépticas, minería de datos, desequilibrados, monitorización, clasificador, API REST ...

Abstract

Epilepsy is a disease whose symptoms are violent seizures and fainting. In that's situations is necessary to apply first aid, but, overnight is difficult to detect and treat in time these crisis. That can cause serious damage to the patients.

In this paper is researched, based on real data, methods which can detect seizures in real time in such a way that caregivers can supply first aids as soon as possible.

In addition, it has been created a REST API and a web page with real time management of beds with the patient's current state and the seizure probability.

Keywords

Epileptic seizures, data mining, unbalanced, monitoring, clasificator, API REST.

Índice general

Índice general	III
Índice de figuras	V
Índice de tablas	VI
Introducción	1
1.1. Material adjunto	1
Objetivos del proyecto	3
Conceptos teóricos	5
3.1. Secciones	5
3.2. Referencias	5
3.3. Imágenes	6
3.4. Listas de ítems	6
3.5. Tablas	7
Técnicas y herramientas	9
4.1. Investigación	9
4.2. Servicio web	10
4.3. Herramientas generales	11
Aspectos relevantes del desarrollo del proyecto	13
Trabajos relacionados	15
Conclusiones y Líneas de trabajo futuras	17

Índice de figuras

3.1. Autómata para una expresión vacía	6
--	---

Índice de tablas

3.1. Herramientas y tecnologías utilizadas en cada parte del proyecto	8
---	---

Introducción

Gracias al avance de las técnicas y algoritmos de minería de datos, disciplinas no directamente relacionadas con la computación se han ido beneficiando de las ciencias de datos, a modo particular, la medicina está desarrollándose hacia modelos más preventivos gracias a las predicciones que se pueden generar utilizando estos métodos.

Es por este motivo que podemos ver en la literatura científica de los últimos años como se relaciona medicina y ciencia de datos, por ejemplo en estudios de detección de caídas [25] o la motorización del sueño para la prevención de apneas [13]. En este trabajo de fin de grado, el objetivo es la detección de crisis epilépticas adscrito al proyecto homónimo vencedor del concurso universitario *Desafío Universidad Empresa* [4]

Aunque el análisis y detección de crisis epilépticas de manera automática ha sido ampliamente explorada por la comunidad científica esta se ha centrado o en el uso de pulseras inteligentes [19] o el uso de encefalogramas (*EEG*) [11, 14, 26] para la predicción de estos eventos. Por tanto, ante pacientes con diversos problemas de salud que impiden el uso de pulseras y de dispositivos de control de actividad cerebral se realiza este proyecto que enfoca la detección de las crisis epilépticas en el uso de sensores de presión y biométricos en la cama donde duerme el paciente.

1.1. Material adjunto

Junto a esta memoria se incluyen:

- **Cuaderno de investigación** con la evolución y pasos realizados durante la investigación junto con los resultados de cada experimento.

- **Anexos** donde se incluyen:
 - Plan de proyectos
 - Requisitos del sistema
 - Diseño del sistema
 - Manual para el programador
 - Manual para el usuario
- **API REST** en **Python-Flask**
- **Experimentos** en *Jupyter Notebook*

Además se puede acceder a través de internet a la página web en producción [7] y al repositorio GitHub del proyecto [6].

Objetivos del proyecto

El objetivo que se plantea es la creación de un clasificador que detecte cuando un paciente dormido padece de una crisis epiléptica, la interfaz de recopilación de datos y la interfaz por la cual el personal médico encargado de la salud del paciente conozca la situación de crisis.

Conceptos teóricos

En aquellos proyectos que necesiten para su comprensión y desarrollo de unos conceptos teóricos de una determinada materia o de un determinado dominio de conocimiento, debe existir un apartado que sintetice dichos conceptos.

Algunos conceptos teóricos de L^AT_EX¹.

3.1. Secciones

Las secciones se incluyen con el comando `section`.

Subsecciones

Además de secciones tenemos subsecciones.

Subsubsecciones

Y subsecciones.

3.2. Referencias

Las referencias se incluyen en el texto usando `cite [?]`. Para citar webs, artículos o libros `[?]`.

¹Créditos a los proyectos de Álvaro López Cantero: Configurador de Presupuestos y Roberto Izquierdo Amo: PLQuiz

3.3. Imágenes

Se pueden incluir imágenes con los comandos standard de \LaTeX , pero esta plantilla dispone de comandos propios como por ejemplo el siguiente:



Figura 3.1: Autómata para una expresión vacía

3.4. Listas de items

Existen tres posibilidades:

- primer item.
- segundo item.

1. primer item.
2. segundo item.

Primer item más información sobre el primer item.

Segundo item más información sobre el segundo item.

▪

3.5. Tablas

Igualmente se pueden usar los comandos específicos de \LaTeX o bien usar alguno de los comandos de la plantilla.

Herramientas	App	AngularJS	API REST	BD	Memoria
HTML5		X			
CSS3		X			
BOOTSTRAP		X			
JavaScript		X			
AngularJS		X			
Bower		X			
PHP			X		
Karma + Jasmine		X			
Slim framework			X		
Idiorm			X		
Composer			X		
JSON		X	X		
PhpStorm		X	X		
MySQL				X	
PhpMyAdmin				X	
Git + BitBucket		X	X	X	X
MikTeX					X
TeXMaker					X
Astah					X
Balsamiq Mockups		X			
VersionOne		X	X	X	X

Tabla 3.1: Herramientas y tecnologías utilizadas en cada parte del proyecto

Técnicas y herramientas

4.1. Investigación

El proyecto ha sido desarrollado aplicando diversas técnicas de minería de datos siguiendo el proceso KDD [21]. Aunque en esta memoria se explicará superficialmente, el desarrollo completo de la investigación se encuentra en el cuaderno de investigación adjuntado.

Para el proceso del clasificador final el flujo KDD ha sido el siguiente:

1. **Selección** Los datos han sido cedidos por la asociación abulense *PRONISA* que contiene jornadas nocturnas con identificadores de la cama, datos de presiones y datos vitales. Estos datos no están balanceados ya que existe una mayor cantidad de datos del paciente durmiendo que bajo una crisis epiléptica.
2. **Preprocesado** Los datos se han limpiado reduciendo ruidos mediante filtros de señal.
3. **Transformación** Se han generado datos estadísticos para series temporales que optimizaban el valor del área bajo la curva PCR [24] desde los datos preprocesados.
4. **Minería de datos** Se ha aplicado un sistema de clasificación doble, en primer lugar se utiliza un árbol de clasificación muy simple que divide entre las situaciones de despierto (bajas presiones en la cama) y acostado, en esta situación se aplica un clasificador *Random Forest* [3] para determinar si hay crisis o no.

5. **Interpretación** Se ha desplegado una aplicación web con datos en tiempo real para evaluar de manera constante la situación actual del paciente.

Este proceso de investigación se realizó sobre *Python* utilizando el ecosistema de librerías *SciPy* [12], en particular la librería de aprendizaje automático *scikit-learn* [17], de computación *NumPy* [16], de análisis de datos *Pandas* [15] y la de dibujado *Matplotlib* [10]. Además, en algunas partes de la investigación se utilizó *Weka* [9] para estudiar métodos que no se encontraban en *scikit-learn* como el *Random Balance* [5] o *Rotation Forest* [20].

4.2. Servicio web

Backend

Las herramientas utilizadas para programar el servidor han sido las siguientes:

Flask [23] *Microframework* de código abierto (BSD) que ofrece una capa de abstracción muy alta de un servicio web.

Jinja [22] Gestor de plantillas de código abierto (BSD) para Python.

Flask-Socketio [8] Integración del servicio de *sockets*, *Socket.IO* [1], compatible con los *WebSockets*.

Gevent y Eventlet [2, 18] Librerías para el uso de tiempo real, asíncrono de hilos para el uso de *Socket.IO*.

Para la programación del sistema de hilos que distribuyen datos en tiempo real se siguieron los paradigmas de la programación orientada a objetos y de programación funcional. El sistema de rutas siguió las guías del *microframework Flask*.

Algunos patrones de diseño utilizados han sido el *Singleton* para la API así como un *Proxy* entre la interfaz web y la lógica de el API.

Frontend

Para el desarrollo de la parte visible de la aplicación se han usado otra serie de herramientas:

Bootstrap [28] *Framework* de *CSS* de código abierto (MIT) creado por *Twitter* para la creación de aplicaciones web redimensionables.

jQuery [29] *Framework* de *JavaScript* de código abierto (MIT) que simplifica el acceso al *HTML DOM* de la página.

Chart.js Librería de *JavaScript* de código abierto (MIT) para la creación de grafos en *canvas*.

4.3. Herramientas generales

Para el desarrollo general del proyecto se han utilizado las siguientes herramientas según el ámbito al que pertenecen:

Servidor

Nginx servidor web y de proxy reverso ligero de alto rendimiento [31] de código abierto (BSD simplificada).

MariaDB sistema de gestión de bases de datos derivado de *MySQL* [30] de código abierto (GPLv2). Este motor es extremadamente compatible con *MySQL* porque es creado como una bifurcación de esto para garantizar la existencia de este motor bajo GPL.

Proxmox es un entorno de virtualización de servidores [32] de código abierto (AGPL). Su función principal es el despliegue y gestión de máquinas virtuales y contenedores.

Anarchy Arch sistema GNU/Linux derivado de *ArchLinux* [27] sobre el cual se ejecuta todo el servidor, está alojado en una máquina virtual del entorno *Proxmox*.

Miscelánea

- **Jupyter Notebooks:** IDE de programación de *Python* basado en *iPython* de código abierto (BSD).
- **PyCharm Professional:** IDE de programación para *Python* avanzado basado en *IntelliJ*.
- **Visual Studio Code:** Editor de código genérico de código abierto (MIT).
- **Postman:** IDE para la ejecución de request *HTTP*.
- **Selenium:** IDE de pruebas sobre web de código abierto (APACHE).
- **CertBot:** Sistema para la firma *SSL* sobre *HTTP* gratuito de *LetsEncrypt* de código abierto (MPL).
- **Overleaf:** Editor de \LaTeX online para el trabajo colaborativo.
- **TexStudio:** Editor de \LaTeX de código abierto (GPLv2).
- **Dia:** Editor de diagramas genérico de código abierto (GPL).
- **StartUML:** Editor de diagramas UML.
- **Codesketch DB:** Traductor bidireccional de código-diagrama para bases de datos.
- **Filezilla:** Aplicación para la transferencia de ficheros sobre *FTP* y *SFTP* de código abierto (GPLv2).
- **GitHub:** Servicio online de *hosting* para repositorios Git.
- **ZenHub:** Servicio online de integración de herramientas *SCRUM* sobre GitHub.

Aspectos relevantes del desarrollo del proyecto

Este apartado pretende recoger los aspectos más interesantes del desarrollo del proyecto, comentados por los autores del mismo. Debe incluir desde la exposición del ciclo de vida utilizado, hasta los detalles de mayor relevancia de las fases de análisis, diseño e implementación. Se busca que no sea una mera operación de copiar y pegar diagramas y extractos del código fuente, sino que realmente se justifiquen los caminos de solución que se han tomado, especialmente aquellos que no sean triviales. Puede ser el lugar más adecuado para documentar los aspectos más interesantes del diseño y de la implementación, con un mayor hincapié en aspectos tales como el tipo de arquitectura elegido, los índices de las tablas de la base de datos, normalización y desnormalización, distribución en ficheros³, reglas de negocio dentro de las bases de datos (EDVHV GH GDWRV DFWLYDV), aspectos de desarrollo relacionados con el WWW... Este apartado, debe convertirse en el resumen de la experiencia práctica del proyecto, y por sí mismo justifica que la memoria se convierta en un documento útil, fuente de referencia para los autores, los tutores y futuros alumnos.

Trabajos relacionados

Este apartado sería parecido a un estado del arte de una tesis o tesina. En un trabajo final grado no parece obligada su presencia, aunque se puede dejar a juicio del tutor el incluir un pequeño resumen comentado de los trabajos y proyectos ya realizados en el campo del proyecto en curso.

Conclusiones y Líneas de trabajo futuras

Todo proyecto debe incluir las conclusiones que se derivan de su desarrollo. Éstas pueden ser de diferente índole, dependiendo de la tipología del proyecto, pero normalmente van a estar presentes un conjunto de conclusiones relacionadas con los resultados del proyecto y un conjunto de conclusiones técnicas. Además, resulta muy útil realizar un informe crítico indicando cómo se puede mejorar el proyecto, o cómo se puede continuar trabajando en la línea del proyecto realizado.

Bibliografía

- [1] Damien Arrachequesne and Erik Little. Socket.io. <https://socket.io/docs/>, feb 2018.
- [2] Karl J Aström. Event based control. In *Analysis and design of nonlinear control systems*, pages 127–147. Springer, 2008.
- [3] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [4] Radio Amiga Burgos. Álar Arnaiz González y José Francisco Díez Pastor – Premiadados del Concurso Desafío Universidad Empresa, Marzo 2018.
- [5] José F Díez-Pastor, Juan J Rodríguez, César García-Osorio, and Ludmila I Kuncheva. Random balance: ensembles of variable priors classifiers for imbalanced data. *Knowledge-Based Systems*, 85:96–111, 2015.
- [6] José Luis Garrido Labrador. Smartbeds - repositorio de código. <https://github.com/joselucross/Smartbeds/>, 2019. [Online; accessed 20-May-2019].
- [7] José Luis Garrido Labrador. SmartBeds - Web en producción. <https://ubu.joselucross.com/>, 2019. [Online; accessed 20-May-2019].
- [8] Miguel Grinberg. Flask-socketio. <https://flask-socketio.readthedocs.io/>, jan 2019.
- [9] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The WEKA data mining software: an update. *SIGKDD Explorations*, 11(1):10–18, 2009.

- [10] John D Hunter. Matplotlib: A 2d graphics environment. *Computing in science & engineering*, 9(3):90–95, 2007.
- [11] Jesper Jeppesen, Sándor Beniczky, A Fuglsang Frederiksen, Per Sidenius, and Peter Johansen. Modified automatic r-peak detection algorithm for patients with epilepsy using a portable electrocardiogram recorder. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4082–4085. IEEE, 2017.
- [12] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. [Online; accessed 07/02/2019].
- [13] Juha M Kortelainen, Mark Van Gils, and Juha Pärkkä. Multichannel bed pressure sensor for sleep monitoring. In *2012 Computing in Cardiology*, pages 313–316. IEEE, 2012.
- [14] Yatindra Kumar, ML Dewal, and RS Anand. Epileptic seizures detection in eeg using dwt-based apen and artificial neural network. *Signal, Image and Video Processing*, 8(7):1323–1334, 2014.
- [15] Wes McKinney et al. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX, 2010.
- [16] Travis E Oliphant. *A guide to NumPy*, volume 1. Trelgol Publishing USA, 2006.
- [17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [18] Daniel Pope. Gevent: asynchronous i/o made easy. EuroPython, 2014. <https://doi.org/10.5446/19958> Lastaccessed : 16May2019.
- [19] Sriram Ramgopal, Sigride Thome-Souza, Michele Jackson, Navah Ester Kadish, Iván Sánchez Fernández, Jacquelyn Klehm, William Bosl, Claus Reinsberger, Steven Schachter, and Tobias Loddenkemper. Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy. *Epilepsy & behavior*, 37:291–307, 2014.
- [20] Juan José Rodríguez, Ludmila I Kuncheva, and Carlos J Alonso. Rotation forest: A new classifier ensemble method. *IEEE transactions on pattern analysis and machine intelligence*, 28(10):1619–1630, 2006.

- [21] Juan José Rodríguez. Introducción a la minería de datos.
- [22] Armin Ronacher, David Lord, Adrian Mönnich, and Markus Unterwaditzer. Welcome to jinja2. <http://jinja.pocoo.org/docs/2.10/>, 2008.
- [23] Armin Ronacher, David Lord, Adrian Mönnich, and Markus Unterwaditzer. Welcome to flask. <http://flask.pocoo.org/docs/1.0/>, 2010.
- [24] Takaya Saito and Marc Rehmsmeier. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3):e0118432, 2015.
- [25] Marie Tolkiehn, Louis Atallah, Benny Lo, and Guang-Zhong Yang. Direction sensitive fall detection using a triaxial accelerometer and a barometric pressure sensor. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 369–372. IEEE, 2011.
- [26] Alexandros T Tzallas, Markos G Tsipouras, Dimitrios G Tsalikakis, Evaggelos C Karvounis, Loukas Astrakas, Spiros Konitsiotis, and Margaret Tzaphlidou. Automated epileptic seizure detection methods: a review study. In *Epilepsy-histological, electroencephalographic and psychological aspects*. IntechOpen, 2012.
- [27] Wikipedia contributors. Arch linux — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Arch_Linux&oldid=896735398, 2019. [Online; accessed 16-May-2019].
- [28] Wikipedia contributors. Bootstrap (front-end framework) — Wikipedia, the free encyclopedia. [https://en.wikipedia.org/w/index.php?title=Bootstrap_\(front-end_framework\)&oldid=895052706](https://en.wikipedia.org/w/index.php?title=Bootstrap_(front-end_framework)&oldid=895052706), 2019. [Online; accessed 16-May-2019].
- [29] Wikipedia contributors. JQuery — Wikipedia, the free encyclopedia. <https://en.wikipedia.org/w/index.php?title=JQuery&oldid=896325154>, 2019. [Online; accessed 16-May-2019].
- [30] Wikipedia contributors. Mariadb — Wikipedia, the free encyclopedia. <https://en.wikipedia.org/w/index.php?title=MariaDB&oldid=897072367>, 2019. [Online; accessed 16-May-2019].

- [31] Wikipedia contributors. Nginx — Wikipedia, the free encyclopedia. <https://en.wikipedia.org/w/index.php?title=Nginx&oldid=896866646>, 2019. [Online; accessed 16-May-2019].
- [32] Wikipedia contributors. Proxmox virtual environment — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Proxmox_Virtual_Environment&oldid=896960348, 2019. [Online; accessed 16-May-2019].