



ESCUELA  
POLITÉCNICA  
NACIONAL



**ESCUELA POLITÉCNICA NACIONAL**

**FACULTAD DE CIENCIAS**



**LIBRO DE RESÚMENES**

**ENERO 2022**

# Contenido

<b>Bienvenida</b>	<b>2</b>
<b>Comité</b>	<b>3</b>
<b>Comité Organizador</b>	<b>3</b>
<b>Comité Científico</b>	<b>3</b>
<b>Colaboradores</b>	<b>3</b>
<b>Contribuciones Orales</b>	<b>5</b>
<b>Ciberseguridad</b>	<b>5</b>
<b>Ciencias Sociales</b>	<b>8</b>
<b>Economía y Finanzas</b>	<b>9</b>
<b>Educación</b>	<b>12</b>
<b>Medio Ambiente</b>	<b>14</b>
<b>Seguridad Ciudadana y otras aplicaciones</b>	<b>17</b>
<b>Posters</b>	<b>20</b>
<b>Ciencias Sociales</b>	<b>20</b>
<b>Ciberseguridad</b>	<b>21</b>
<b>Economía y Finanzas</b>	<b>23</b>
<b>Medio Ambiente</b>	<b>25</b>
<b>Negocios e Industria</b>	<b>31</b>
<b>Seguridad Ciudadana y Otras Aplicaciones</b>	<b>32</b>
<b>Educación</b>	<b>34</b>
<b>Cronograma</b>	<b>36</b>

## Bienvenida

Estimados participantes,

En nombre del comité organizador, me gustaría expresar nuestro agradecimiento por compartir y discutir los resultados de su investigación actual en el marco de la Ciencia de Datos, Estamos convencidos de que estos espacios potencian y fomentan las posibilidades de establecer y fortalecer redes de trabajo y amistad, son también el lugar donde nacen nuevas ideas y proyectos y, finalmente, juegan un papel clave en la motivación de jóvenes investigadores y estudiantes.

El Congreso de Investigación Aplicada a Ciencia de Datos y II Congreso Nacional de R Users Group-Ecuador, da inicio a una nueva etapa el cual se enfoca en conocer los aspectos fundamentales, metodológicos y prácticos de los métodos de matemática aplicada utilizados en los diferentes campos de las Ciencias de Datos. Además, de su implementación en la práctica a través de la implementación de metodologías utilizando el software estadístico de programación R.

La pandemia provocada por el COVID-19 obligó, lamentablemente, a cambiar el formato establecido de nuestros congresos a modalidad virtual. Pero por otro lado, esto también proporcionó ventajas. Una de estas ventajas es que se puede contar con la participación de 9 destacados expositores, de los cuales 7 son internacionales de países como España, Chile, Argentina y Colombia y 2 son de Ecuador, quienes son líderes reconocidos en sus campos sin tener que incurrir en mayores problemas logísticos.

Adicionalmente, fue posible recibir aportes de investigadores, como se mencionó anteriormente, el programa incluye 9 expositores magistrales, 17 contribuciones orales y 20 poster. Previo al congreso, hemos organizado un ciclo de webinar con los expositores magistrales llevado a cabo del 29 de noviembre al 01 de diciembre de 2021. Este evento solo es posible gracias al esfuerzo del comité organizador del congreso, y gracias a la ayuda comprometida de los estudiantes, colaboradores del CCDE, asistentes y personal administrativo del Departamento y de la EPN. Por supuesto, sin el compromiso de los investigadores y el público, esto no habría sido posible ni tendría un propósito.

¡Gracias!

¡Esperamos que el Congreso de Investigación Aplicada a Ciencia de Datos y II Congreso Nacional de R Users Group cumpla con sus expectativas y les deseamos una fructífera e interesante conferencia!

Miguel Flores Coordinador del Congreso

## Comité

### Comité Organizador

- Dr. Miguel Flores - Coordinador del Congreso
- Dr. Christian Vásconez - Miembro de la Comisión del Congreso
- Dr. Edison Loza - Miembro de la Comisión del Congreso
- MSC. Yandira Cuvero - Miembro de la Comisión del Congreso
- Ing. Ana Belén cabezas - Miembro de la Comisión del Congreso
- Ing. Lisette Zambrano - Miembro de la Comisión del Congreso

### Comité Científico

Salvador Naya Fernández  
 Universidad Da Coruña, España  
[salvador.naya@udc.es](mailto:salvador.naya@udc.es)  
[http://dm.udc.es/staff/salvador\\_naya/](http://dm.udc.es/staff/salvador_naya/)

Javier Tarrío Saavedra  
 Universidad Da Coruña, España  
[Javier.tarrio@udc.es](mailto:Javier.tarrio@udc.es)  
<https://pdi.udc.es/es/File/Pdi/JK2GF>

Jorge Mateu  
 Universitat Jaume I, Castellon, España  
 mateu@uji.es  
<https://www3.uji.es/~mateu/>

Jorge Antonio Párraga Álava Ph.D.  
 Universidad Técnica de Manabí, Ecuador  
[jorge.parraga@utm.edu.ec](mailto:jorge.parraga@utm.edu.ec)

Adriana Uquillas Andrade Ph.D.  
 Escuela Politécnica Nacional, Ecuador  
[adriana.uquillas@epn.edu.ec](mailto:adriana.uquillas@epn.edu.ec)

Carlos Almeida Ph.D.  
 Escuela Politécnica Nacional, Ecuador  
[carlos.almeidar@epn.edu.ec](mailto:carlos.almeidar@epn.edu.ec)

### Colaboradores

- Alexander Constante
- Alexandra Maigua
- Andrés Vinueza

- Cristian Solorzano
- Christopher Aguirre
- Danilo Loza
- Diego Heredia
- Felipe Guevara
- Génesis Moreno
- Héctor Guacapiña
- Jeysson Chuquin
- Juan Játiva
- Kathya Gavilanes
- Omar Llambo
- Priscila Guayasamín
- Yasmina Vizuite

## Contribuciones Orales

### Ciberseguridad

#### Detección Automática de Noticias Falsas en Español: Sátira Política Ecuatoriana

Ing. Nicolás Mafla

##### Biografía

Nacido en Quito, cursó sus estudios secundarios en el colegio Giovanni Antonio Farina del cantón Rumiñahui, graduándose como bachiller en Ciencias en el año 2014. Para el 2015 ingresa a la carrera de Ingeniería Matemática con mención Estadística e Investigación de Operaciones en la Escuela Politécnica Nacional. Durante este periodo formó parte del proyecto CLAVEMAT como tutor voluntario, brindando apoyo académico en materias fundamentales para estudiantes de niveles iniciales de ingeniería. Obtuvo su título como Ingeniero matemático a mediados del año 2021 y actualmente trabaja en Altscore, una fintech ecuatoriana, como científico de datos.

##### Resumen

La circulación de noticias falsas en internet, especialmente las de sátira política a través de redes sociales, ha afectado a la mayoría de la población ecuatoriana. Este trabajo presenta una metodología basada en el aprendizaje estadístico que detecta de forma precisa y automática noticias falsas en español utilizando técnicas de aprendizaje automático y procesamiento del lenguaje natural. El documento comienza presentando conceptos básicos relacionados con las noticias falsas y trabajos relacionados con su detección automática. La segunda sección explica el proceso de creación del corpus de noticias, procesamiento de los textos, representación numérica con TF-IDF y entrenamiento de algoritmos de clasificación supervisados con dos conjuntos de datos diferentes. Los resultados obtenidos del entrenamiento se analizan en la tercera sección, siendo los modelos con máquinas de soporte vectorial los que ofrecen mejores predicciones. Finalmente, las conclusiones de la investigación y el trabajo futuro se presentan en la cuarta sección.

#### Analysis and Design of a Predictive Model for Phishing Detection Based on Url and Email Corpus

Srta. Fernanda Albán Toapanta

##### Biografía

Nacida en Latacunga, sus estudios primarios y secundarios los realizó en esta ciudad; para continuar con su formación académica, cambia de residencia a la capital de los ecuatorianos

donde se encuentra las instalaciones de la Escuela Politécnica Nacional, en su trayectoria en esta institución realizó sus prácticas preprofesionales en Petroamazonas en el área de RR.HH., donde realizó un modelo para la mejora de la productividad en el personal de campo. En la actualidad es egresada en la carrera de Ingeniería Matemática mención Estadística e Investigación Operativa. Dictó cursos en Arcotel, además, realizó análisis estadísticos para Jedral in software & AI, we trust y para Sercop. Para su trabajo de titulación se enfocó en el área de Cyber security en forma específica la detección de Phishing.

## Resumen

Uno de los delitos cibernéticos más reportados a nivel mundial es el phishing, por esta razón, actualmente se está desarrollando diversos sistemas anti-phishing (APS) para identificar este ataque en sistemas de comunicación en línea. A pesar de los esfuerzos, este ataque continúa sin cesar, teniendo como causas, la detección errónea en el ataque de día cero, el alto costo computacional y las tasas altas de falsificación. Aunque el enfoque de Machine Learning (ML) ha logrado una tasa de precisión favorable, se debe considerar que la elección y el rendimiento del vector de características es un punto clave para obtener un nivel de precisión elevado. En este trabajo, proponemos un modelo predictivo basado en ML y analizar la eficiencia de algunos esquemas anti-phishing que sirvieron para entender esta temática. El modelo propuesto consta de un módulo de selección de características que se utiliza para la construcción del vector final. Estas características se extraen de la URL, las propiedades de la página web y del corpus de correo electrónico, utilizando un sistema basado en componentes incrementales para presentar el vector resultante. El sistema utiliza modelos de clasificación, Random Forest y Naïve Bayes, que han sido entrenados en el vector de rasgos. Los experimentos se basaron en dataset compuestas por instancias de phishing y benignas. Utilizando la validación cruzada, los resultados experimentales indican una precisión del 97,5% para las bases mencionadas en otros trabajos, mientras que para el abordaje de esta investigación a nivel local se obtuvo una precisión del 96,5%

## A security prototype for improving home security through LoRaWAN technology

Ing. Miguel Parra  
 Ing. Edwin Ávila  
 MSc. Jhonattan J. Barriga  
 PhD. Sang Guun Yoo Park  
 MSc. Roberto Andrade

## Biografía

### Ing. Miguel Parra

Ing. en Sistemas Informáticos y de Computación de la Facultad de Ingeniería en Sistemas de la Escuela Politécnica Nacional. Trabaja como desarrollador full stack en Finanzmotoros. Apasionado por la investigación en tópicos concernientes a desarrollador web y criptotrading.

### **Ing. Edwin Ávila**

Ing. en Sistemas Informáticos y de Computación de la Facultad de Ingeniería en Sistemas de la Escuela Politécnica Nacional. Trabaja como desarrollador full stack en Veris. Sus intereses están centrados en el campo de la ciberseguridad y pentesting.

### **MSc. Jhonattan J. Barriga**

Estudiante de doctorado de Sistemas de Seguridad en la Facultad de Ingeniería en Sistemas de la Escuela Politécnica Nacional. Obtuvos su grado de MSc. (Distinction) Computer Forensics and Systems Security por la Universidad de Greenwich, Londres, Inglaterra (2011). Ing. En Sistemas de la Escuela Politécnica del Ejército (2007). Profesor a tiempo completo del Departamento de Informática y Ciencias de la Computación. Sus principales intereses son: Malware, Penetration Testing, IoT Security, Informática Forense. Ha trabajado como IT Architect para IBM del Ecuador (2007 – 2010). Para TATA Consultancy Services (2012 – 2015) como Information Security Coordinator. Hacker Ético certificado por Ec-Council (CEH).

### **PhD. Sang Guun Yoo Park**

Profesor a tiempo completo del Departamento de Informática y Ciencias de la Computación. Obtuvo su título de PhD del Departamento de Ciencias de la Computación e Ingeniería de la Universidad de Sogang, Seoul, Corea en 2013. Co-fundador de Extremo Software (Microsoft Gold Certified Partner) donde trabajó como CTO desde el 2001 al 2005. Del 2005 al 2007 colaboró como profesor del Departamento de Ciencias de la Computación y Multimedia por la Universidad Internacional del Ecuador. Del 2006 al 2007 como consultor de IT para la SENAIN. Se ha desempeñado como Jefe de Investigación e Ingeniería para LG Electronics de la República de Corea del Sur. Actualmente es un punto de contacto Nacional de la comisión Europea H2020. Miembro Senior IEEE y SCIEI.

### **MSc. Roberto Andrade**

Estudiante de doctorado en Sistemas de Seguridad en la Facultad de Ingeniería de Sistemas en la Escuela Politécnica Nacional (EPN), su maestría es en Gestión de Redes y Telecomunicaciones en la Escuela Politécnica del Ejército en 2013 y su título de ingeniero es en Electrónica y Telecomunicaciones en la Escuela Politécnica Nacional (EPN) en 2007. Oficial de Seguridad del Ministerio de Educación de Ecuador (MINEDUC) en 2015, Coordinador de Infraestructura Tecnológica en la Secretaría Nacional de Planificación SENPLADES 2013-2014, Centro de datos, seguridad y administración de redes en SENPLADES y Tecnología Sucre 2009-2013 e Ingeniería Técnica para sistemas VoIP en SERATVoIP 2007-2011. Es instructor técnico certificado de CCNA, CCNP y CCNA Security en EPN desde 2010 hasta la fecha.

### **Resumen**

Wireless networks' constant expansion and the growth of smart devices connectivity have motivated the search for new communication solutions to provide benefits for social problems like citizen insecurity, where home robbery is one of the most relevant. A solution to this type



of problem is home automation, by using sensors that can detect intrusions at a house. This work proposes the development of an LPWAN prototype by using LoRa technology for detecting intruders in houses in a residential area. This approach focuses on integrating LoRaWAN protocol combined with MQTT and an API REST to generate notification alerts for the security personnel and the house owner. A mobile application is used by the house owner to handle nodes. A web application has been developed as well so that security personnel could manage user authentication and monitor notification alerts generated by the nodes deployed in different houses. Push notifications have been enabled whenever an intrusion occurs, or a node is disconnected.

## Ciencias Sociales

### Construction of a spatial inequality social indicator by Kernel Nonlinear Principal Components Analysis, using the Survey of Living Conditions 2014

Ing. Jared Abigail Valencia

#### Biografía

Nacida en Quito, graduada del Colegio Particular Jim Irwin como Bachiller en Ciencias. En el 2015 ingresa a la carrera de Ingeniería Matemática, mención Estadística e Investigación Operativa, de la Escuela Politécnica Nacional, donde durante su estancia formó parte del proyecto CLAVEMAT como voluntaria, brindando apoyo académico en materias fundamentales de la carrera; también dictó cursos de programación en Matlab en la ESPE - Universidad de las Fuerzas Armadas.

Además, fue vicepresidenta financiera de la Asociación de matemática e ingeniería matemática de la EPN; obtuvo su título como Ingeniera matemática en junio del 2021 con su proyecto enfocado en el análisis espacial de la desigualdad social en el Distrito Metropolitano de Quito.

Actualmente es Analista de Negocios en el área de Medios de Pago de Produbanco – Grupo Promerica y cursa el máster en Inteligencia de Negocios en la Universidad Internacional de la Rioja.

#### Resumen

El método de análisis de componentes principales basado en un kernel (KPCA) es una generalización de los métodos PCA no lineales usando kernels definidos positivos. El objetivo es encontrar variables proyectadas en un kernel inducido, con máxima varianza, del espacio de características. Además, si se utiliza una función de decaimiento de distancia como kernel, el KPCA puede considerar la interacción espacial en los datos.

El KPCA ha sido utilizado para simulaciones de crecimiento urbano, reconocimiento de patrones, reconocimiento facial, monitoreo de procesos dinámicos no lineales, predicción de series de tiempo, aprendizaje y reconocimiento de emociones, entre otros. Pero ¿qué hay acerca de la desigualdad social?

La desigualdad social ocurre cuando un grupo de personas recibe un tratamiento diferente como consecuencia de su posición social, situación económica, género, la religión que profesan entre otros aspectos. Todos estos factores se intersecan en los territorios, por lo que varios estudios sugieren que, el lugar donde una persona nace o vive influye en su desarrollo social, así como el acceso y distribución adecuada de oportunidades de desarrollo, dando como resultado desigualdades territoriales.

Entonces, el objetivo es analizar las variables de la Encuesta de Condiciones de Vida 2014, Sexta Ronda, para la construcción de un índice de desigualdad social que considere la influencia espacial y geográfica de los datos para medir la desigualdad social en Ecuador, específicamente en el Distrito Metropolitano de Quito, usando KPCS con una función de decaimiento de distancia como kernel.

Como resultado, el estudio mostró que existían muchas regiones de Quito que presentan desigualdad social alta en la región periférica de la ciudad. De hecho, esto no es nuevo ya que esta tendencia de menores oportunidades en los lugares periféricos ocurre alrededor del mundo.

Una observación interesante fue que el 75% de regiones analizadas presentó más del 70% de desigualdad social y el 7% presenta menos del 14%. Se compararon estos resultados con un estudio de necesidades básicas insatisfechas realizado en el Distrito Metropolitano de Quito y, en efecto, coinciden.

El método KPCA generó una representación de la realidad social de Quito tomando en cuenta la interacción espacial. Es evidente que la desigualdad social en el DMQ es un problema no resuelto, y las autoridades correspondientes deberían proponer nuevas políticas de desarrollo e inclusión para mejorar la calidad de vida de los ciudadanos del DMQ.

## Economía y Finanzas

### Soluciones analíticas en banca mediante algoritmos de ciencia de datos

Jorge Martín Arevalillo, PhD

#### Biografía

Licenciado en ciencias matemáticas, premio extraordinario fin de carrera, y doctor en ciencias matemáticas por la Universidad Nacional de Educación a Distancia (UNED) con más de veinte años de experiencia docente, investigadora y de transferencia de conocimiento en Ciencia de Datos tanto en Europa como en América. Actualmente desarrolla actividad docente e investigadora como profesor del departamento de Estadística e Investigación Operativa de la UNED. Ha colaborado en proyectos de asesoramiento científico-técnico y transferencia de conocimiento y en proyectos de investigación con BBVA Perú, McDonald's España, Yell Publicidad, Mutua Madrileña Automovilista, Biomedica Molecular Medicine, Hospital Universitario La Paz, Salford Systems (USA) y University of Maryland School of Medicine (USA).

También ha dictado cursos y programas de formación en Ciencia de Datos para instituciones públicas y privadas: Data Mining Consulting (Perú), Decide Soluciones, Boston Consulting Group, Centro Internacional de Formación Financiera y Universidad de Cádiz. Sus líneas de investigación tienen que ver con el desarrollo y aplicación de algoritmos para el aprendizaje y la extracción de conocimiento de los datos; abarcan diversas disciplinas con contribuciones científicas en revistas de Matemáticas, Estadística, Computación, Inteligencia Artificial y Biomedicina. El profesor Jorge Martín Arevalillo es ponente habitual en congresos internacionales sobre Big Data y Ciencia de Datos. Su trabajo ha tenido eco en medios de comunicación como Nova Ciencia y Cataluña Económica. La trayectoria académica e investigadora del Dr. Martín Arevalillo ha sido reconocida con el premio UNED – BANCO SANTANDER de investigación, transferencia y divulgación 2018.

## Resumen

El proceso transformativo del mundo en que vivimos hacia la digitalización ha hecho emerger una nueva visión del negocio y la gestión empresarial basada en el Big Data y las tecnologías que lo soportan como palancas de cambio para la eficiencia y la mejora competitiva. Se examinará el fenómeno de la explosión del dato, que denominaremos el big bang de los datos (big bang data), haciendo hincapié en el papel protagonista que juega la Ciencia de Datos ante el desafío de transformar el dato en valor mediante el uso de soluciones analíticas en una industria tradicional como la banca. Este desafío abre un escenario de oportunidades tanto para la investigación académica en Ciencia de Datos como para la transferencia de los resultados de la investigación mediante el desarrollo y aplicación de algoritmos para la analítica de negocio y la gestión data – driven de la operativa comercial en banca. Se examinan algunos proyectos desarrollados en banca como casos de uso ilustrativos de una propuesta de transferencia que genera valor de negocio. Los casos presentados forman parte de las iniciativas de transferencia en Ciencia de Datos articuladas desde la Oficina de Transferencia de Resultados de Investigación (OTRI) de la Universidad Nacional de Educación a Distancia (UNED) para todas aquellas empresas e instituciones interesadas en adoptar soluciones analíticas de valor basadas en la investigación.

## El efecto asimétrico del precio del petróleo en la economía ecuatoriana

Paul Carrillo-Maldonado

## Biografía

Ingeniero en Ciencias Económicas y Financieras de la Escuela Politécnica Nacional (EPN) y Licenciado de la Universidad Jean Monnet (convenio con la EPN). Obtuvo su Maestría en Economía con mención en Economía del Desarrollo (FLACSO). Es candidato a Ph.D. en Economía del Desarrollo en el Programa Doctoral de FLACSO. Ha trabajado para instituciones públicas del Ecuador (BCE, SRI y MCPE), así como para organismos multilaterales (BID, CEPAL, OEI y BM). Ha publicado artículos en revistas académicas nacionales e internacionales.

## Resumen

El objetivo del presente estudio es identificar si el precio internacional del petróleo tiene efectos asimétricos en el crecimiento económico del Ecuador entre 2003 y 2019. Para observar esta no-linealidad en la economía, primero, se separa la evolución del precio del crudo en dos estados: variaciones positivas y negativas, y, posteriormente, se estima el efecto del precio del petróleo en el producto interno bruto a través del modelo de proyecciones locales. Los resultados muestran que las variaciones negativas (decrecimiento) del precio del petróleo tiene mayor efecto en la economía ecuatoriana que el incremento del mismo precio internacional.

## Credit Scoring: Aplicando Técnicas de Regresión Logística y Modelos Aditivos Generalizados para una Cartera de Crédito en una entidad financiera

Ing. Andrés Suquillo

## Biografía

Graduado del Colegio Nacional Juan Pío Montúfar con especialidad Físico Matemático. Ingeniero Matemático, mención Estadística e Investigación Operativa, de la Escuela Politécnica Nacional. Especializado en estadística y ciencia de datos.

Actualmente se desempeña como Especialista de Modelamiento de Fraudes en Banco Pichincha y como consultor independiente en desarrollo de aplicativos webs, motores analíticos y de monitoreo. Adicionalmente trabaja como profesor independiente dictando cursos relacionados a modelos estadísticos y ciencia de datos. Cuenta con experiencia en implementación de modelos estadísticos y automatización de procesos mediante programación en R y Python.

## Resumen

El presente proyecto tiene como finalidad el estudio de una metodología estadística basada en medidas de asociación, medidas de separación y presentar el Modelo Aditivo Generalizado como una alternativa prometedora a la Regresión Logística, los cuales se aplican en la construcción de un modelo estadístico, que permita estimar la probabilidad de incumplimiento (PD) de una cartera de algún tipo de crédito concedido.

La regresión logística es el modelo estadístico más utilizado en la industria de calificación crediticia. A pesar de sus ventajas en la fácil interpretación y el bajo costo computacional, la regresión logística está bajo la crítica de la dificultad de modelar las características no lineales del efecto de los predictores sobre la variable dependiente y, por lo tanto, podría dar lugar a resultados insatisfactorios.

En el presente estudio se plantea la utilización de una técnica conocida como Modelo Aditivo Generalizado introducido por Hastie y Tibshirani (1990), que proporciona la capacidad de detectar la relación no lineal y no monótona entre la variable dependiente y los predictores sin sacrificar la interpretabilidad.

El rendimiento de los modelos se evalúa y comparan utilizando el valor del estadístico de Kolmogórov - Smirnov (KS), el área bajo la curva de Características operativas del receptor (ROC) y el test de GINI.

Adicionalmente, haciendo uso de software estadístico R, se implementan las técnicas descritas, el cual permite obtener los resultados de manera automática disminuyendo considerablemente el tiempo empleado.

## Educación

### Evaluación bibliométrica de los antimicrobianos y las contribuciones en WOS (2000-2021)

Iván Analuisa Aroca

#### Biografía

Iván Alberto Analuisa Aroca, de profesión Ingeniero Zootecnista, con maestría en Administración de Empresas mención Proyectos graduado en la Escuela Superior Politécnica de Chimborazo, me encuentro realizando el proceso de defensa de tesis de la maestría de Investigación en Estadística Aplicada de la Universidad Técnica de Manabí, además de realizar mis estudios doctorales en la Universidad de Córdoba actualmente cursando el Tercer año. He sido docente por cerca de 10 años en varias instituciones de educación superior, en pregrado y postgrado. Además, Coordinador de Carrera de Administración de Empresas y Director de Vinculación con la Sociedad. Reconocido por la institución por el trabajo realizado en los diferentes proyectos sociales integrales a cargo. Participante de cursos de actualización, ponencias en congresos nacionales e internacionales, y capacitador de varios programas nacionales. Formo parte del proyecto de investigación cadenas de valor del maíz en Ecuador. Formo parte de las Redes de Investigación “REDUS” y “Encuentra tu par

#### Resumen

Este estudio ofrece una descripción bibliométrica de la literatura sobre los antimicrobianos en diferentes áreas. Se analizaron y discutieron el número de publicaciones anuales, principales autores, países, organizaciones productivas, las coautorías y colaboración internacional donde participan los autores. El objetivo analizar las publicaciones sobre los antimicrobianos en la base de datos bibliográfica “WOS” Metodología: Se utilizaron palabras clave para la búsqueda “antimicrobials” e “ingredientes alimenticios” utilizando la base de datos Web of Science. Se presentaron informes cualitativos y cuantitativos recuperados utilizando mapas visuales de la aplicación Vosviewer. Resultados: Se consideraron un total de 2714 publicaciones de revistas. El creciente apoyo de las organizaciones financieras para las publicaciones demuestra un incremento desde el 2011 hasta la fecha. Los documentos estudiados fueron publicados en 103 países, concentrándose en áreas relacionadas con la Tecnología de Ciencia de los alimentos y Química con 1860 y 1718 respectivamente. El idioma inglés ocupa el primer lugar con 2635

publicaciones, muy por detrás están el idioma coreano (53) y español (40). Conclusiones: El incremento en las publicaciones durante la última década a nivel mundial, debe fomentar la colaboración internacional, esperando que el estudio bibliométrico ayude a los planificadores en política pública fomentar el desarrollo e investigación en temáticas relacionadas.

## **Retos en la reutilización de datos para la lucha anticorrupción, un estudio descriptivo desde la Sociedad Civil**

Susana Cadena Ph.D.

### **Biografía**

Doctora en Informática, Docente- investigador Universidad Central del Ecuador, Co-Fundador y Directora en Tecnología e Investigación de Fundación Datalat. Docente Investigadora con más de 20 años de experiencia en investigaciones y docencia en pregrado y posgrado en las áreas de: Gestión de Datos, Auditoría Informática y Tecnologías para la Educación. Ha desempeñado varios cargos de gestión con impacto nacional. Ha realizado varios proyectos con diferentes países en proyectos relacionados con soluciones tecnológicas y gestión de datos. Actualmente se encuentra participando en la elaboración de propuestas estratégicas como parte de las redes: REDAM (Red Ecuatoriana de Datos Abiertos y Metadatos), RICHSL (Red de Conocimiento de Software y Hardware Libre). Red CUDI- México (Corporación Ecuatoriana para el desarrollo de Internet, Grupo Gobierno de TI).

### **Resumen**

La apertura de datos en formatos reutilizables es un proceso que permite ejercer los derechos digitales en un entorno cada vez más digital, esta tendencia que nació en el año 2009, y se ha difundido en todo el mundo permitiendo inclusive la generación de nuevos modelos de negocio. La apertura de datos es solo el inicio de un proceso que permita a la ciudadanía generar valor de este acceso, aquí radica la importancia de capacidades de la ciudadanía en análisis de datos, inteligencia artificial, machine learning entre otras, es decir conocimientos relacionados a ciencia de datos. La importancia de la apertura de los datos radica en la capacidad de los ciudadanos de generar nuevos modelos que permitan mejorar la política pública y el bien común de la ciudadanía. En el Ecuador la política y la guía de datos abiertos fue publicada en el año 2020, pero ¿qué avances ha tenido esta política?, ¿qué se conoce en la ciudadanía acerca del uso de estos datos?, ¿Qué retos se enfrenta en la utilización de los datos?, estas son las preguntas directrices de la presente investigación de tipo descriptivo, que permitió establecer las características de tienen la guía y la política, la implementación del portal de datos abiertos, el número de conjuntos publicados, un levantamiento de otros portales de datos abiertos en el país, para concluir con los retos que se tienen desde la reutilización de estos datos en temas anticorrupción con el fin de innovar en los procesos de educación formal y no formal.

## **Modelo de predicción del rendimiento académico para el curso de nivelación de la Escuela Politécnica Nacional a partir de un modelo de aprendizaje supervisado**

Ing. Karen Calva

### **Biografía**

Karen Priscilla Calva Yaguana, su campo de especialización es la ciencia de datos enfocada a alcanzar metas empresariales usando Cloud Computing. Es Ingeniera Matemática graduada de la Escuela Politécnica Nacional con postgrado en Dirección Logística por la Universidad Internacional de la Rioja y Diplomado en liderazgo por parte del INCAE. Actualmente se desempeña como jefa de Analítica de datos en Kushki y como docente en la Sociedad Ecuatoriana de Estadística en temas de geo-estadística, gestión de procesos y logística en levantamiento de información, desarrollo de aplicativos webs con motores analíticos y aprendizaje automático.

### **Resumen**

En el presente artículo se aplica un modelo de aprendizaje automático supervisado que predice la probabilidad de que un estudiante de la Escuela Politécnica Nacional apruebe el curso de nivelación. Para llevar a cabo esta tarea se describe una metodología estadística basada en gradient boosting y regresión logística donde el problema de aprendizaje se formula en términos de la minimización de la función de error mediante el método del descenso del gradiente. Para explicar la probabilidad de aprobación se toman en consideración dimensiones sugeridas por la literatura relacionadas a variables socioeconómicas, demográficas, familiares, institucionales y de desempeño académico en la postulación y en el curso de nivelación que tiene el estudiante. Los resultados del modelo de árbol de decisión muestran un nivel de precisión del 96% en el conjunto de datos de prueba, con un área bajo la curva ROC de 89.1, siendo estos niveles generalmente aceptados. Por otro lado, los resultados de la regresión logística sugieren que factores como la calificación ponderada del primer bimestre, la calificación con la que postuló, su jornada de estudios, su ubicación geográfica de origen, entre otras, afectan de una u otra manera a la probabilidad del estudiante, de aprobar el curso de nivelación.

### **Medio Ambiente**

#### **Hydrosanitary management scene with COVID-19 for the city of Quito - Ecuador, using AI, prediction algorithms and Data Science**

Monica Delgado-Yanez  
Myriam Hernández-Álvarez

## Biografía

Mónica Delgado-Yáñez, ingeniera Civil de la Escuela Politécnica Nacional del Ecuador, Msc en ciencias Sistemas de Información Geográfica, Mgs en Gestión ambiental, MBA en Administración Ambiental y de ecosistemas. PhD. Universidad de Valencia Contaminación, Toxicología y Sanidad Ambiental, biología celular – bioinformática, estudiante de Doctorado EPN sistemas. Fue directora de investigación INAMHI 2014-2017, docente universitaria desde el 2004 UDLA, UISEK. Investigador acreditado SENECYT REG-INV-19-03639

Myriam Hernández-Álvarez Recibió el título de Ingeniera en Electrónica y Telecomunicaciones de la Escuela Politécnica Nacional, Ecuador, el M.Sc. en Ciencias de la Computación de la Universidad de Ohio, Athens, OH, USA, y el Ph.D. en Aplicaciones Informáticas por la Universidad de Alicante, España. Fue Decana de la Escuela de Ingeniería de Sistemas, de 2014 a 2019, y Directora del Programa de Doctorado en Informática de la Escuela Politécnica Nacional, de 2016 a 2019.

## Resumen

In this research, we generate a map for hydrosanitary management to mitigate damages produced by the COVID-19 pandemic in Quito, Ecuador. For this, we used Bio-Computer Systems and Artificial Intelligence algorithms applied over BigData. With these algorithms, we generated health models to predict and prevent collapses as leachate, liquid waste loads, and water requirement overloads. We designed Deep Learning methods using Big Data, numerical resolution of equations of the sanitary model, which includes nesting and temporal discretization. Such models also provide secondary data on increased precipitation likelihood for public interest sectors. We also generated IDF intensity, duration, and frequency curves with an urban runoff coefficient for the rational method in recurrence storms of 5-10, and 0.4-0.75 years; which affect hydro-sanitary overloads in requirements or waste. Our model adequately describes changes in sanitary conditions due to the COVID-19's presence and predicts its future problems. Our model incorporates edge conditions, hydrological considerations, geostatistics, and prediction techniques. Lastly, this paper presents a management map. In sum, this research aims to help improve Ecuador's health issues caused by SARS-CoV2 and COVID-19.

## Adaptación del Método de agrupación K-Medias para datos funcionales Correlacionados espacialmente con aplicación a datos del Índice de Vegetación de diferencia normalizada de los páramos del Ecuador.

Sr. Jeysson Chuquin

Srta. Alexandra Maigua

## Biografía

Jeysson Chuquín y Alexandra Maigua son egresados de la carrera de Ingeniería Matemática de la Facultad de Ciencias de la Escuela Politécnica Nacional. Fueron ponentes en la vigésima primera conferencia anual del European Network for Business and Industrial Statistic (ENBIS-21)



y en el Simposio de Estadística: Modelando la Incertidumbre, evento organizado por el Centro de Estudios e Investigaciones Estadísticas de la Facultad de Ciencias Naturales y Matemáticas de la ESPOL.

Los intereses de investigación de Jeysson incluyen estadística aplicada, modelos de aprendizaje automático, minería y ciencia de datos, geoestadística, control de calidad, modelos de riesgo económico, ethical-hacking y computación de alto rendimiento HPC.

Los intereses de investigación de Alexandra incluyen geoestadística, minería y ciencia de datos, modelos de aprendizaje automático, estadística aplicada.

## Resumen

La dependencia espacial en los datos medioambientales es un criterio influyente en los procesos de agrupación, ya que los resultados obtenidos aportan información relevante. Como los métodos clásicos no tienen en cuenta la dependencia espacial, la consideración de esta estructura produce resultados inesperados y agrupaciones de curvas que pueden no ser similares en forma o comportamiento. En este trabajo se realiza la agrupación mediante el método de k-medias modificado para datos funcionales espacialmente correlacionados aplicado a datos del Índice de Vegetación de Diferencia Normalizada (NDVI) de los páramos del Ecuador. Para ello, se implementan índices de calidad que permiten obtener el número adecuado de grupos. La metodología se base en una ponderación de la matriz de distancia entre las curvas con el trazo-variograma calculado con los coeficientes de las funciones base resultantes de la suavización de los datos. Para la validación del método se realizan algunos escenarios de simulación, así como un caso de aplicación a datos del NDVI; obteniéndose cinco regiones distribuidas latitudinalmente.

## Agriculture in the Digital Era: Present and Challenges

Guillermo Marcillo., PhD.

Mallory K. Vestal., PhD

Ms. Kevyn Thompson

## Biografía

**Guillermo Marcillo., PhD. (Speaker)**

**Assistant Professor, Agricultural Statistics and Analytics**

**West Texas A&M University – Department of Agricultural Sciences**

Dr. Guillermo Marcillo conducts research at the intersection of agriculture, data science, and analytics through synergistic collaborations between University research groups, federal and state agencies, and industry partners. Guillermo is interested in the use of digital technologies: remote sensing, satellite imagery, and other precision agriculture inputs, to apply computational methods (Statistical modeling, Dynamic Simulations) leading to enhancing agricultural decision-making. Dr. Marcillo received a Ph.D. in Crop Physiology (Systems Modeling) from Iowa State University, a Master in Agricultural Economics and Statistics from New Mexico State University,

and was a postdoctoral Data Scientist at the University of Illinois and the Department of Agriculture (USDA).

**Mallory K. Vestal, PhD**

**Associate Professor, Agricultural Business and Economics**

**West Texas A&M University – Department of Agricultural Sciences**

Dr. Vestal teaches courses on Agricultural Economics, Farm and Ranch Management, Futures and Options, and Advanced Farm and Ranch Management. She also conducts research in topics relevant to livestock production economics, marketing, price analysis, policy, and economics of the equine industry. Dr. Vestal received M.S and Ph.D. degrees in Agricultural Economics from Oklahoma State University.

**Ms. Kevyn Thompson**

**Graduate Research Assistant, Agricultural Business and Economics**

**West Texas A&M University – Department of Agricultural Sciences**

Kevyn is originally from New Mexico. Currently, she is a graduate student at West Texas A&M University and is working towards her Master's in Agricultural Business and Economics. Her research is about the impact of water availability on Land Values in the state of Kansas.

## Resumen

The management of high-resolution data requires expertise in statistical and technological areas for an effective use thereof. In a digitized world where data is prevalent in all human activity, Agriculture is no exception. Data Science; understood as a reinterpretation of traditional Statistics from a computational approach, allows agricultural researchers to effectively use the field and out-of-season information to optimize agricultural decisions. Here, I present examples of the integration of digital and precision sources (sensors, spectral information) as well as traditional sources (field experiments, historical observations) to generate supervised predictions (e.g., GAM models, Support Vector Machines) in two areas of agricultural impact: genetic improvement of crops and cost-effective water rights allocation in semi-arid environments. Effective use of new technologies adds value to large volumes of information, but it also poses challenges. It is the task of our universities to train a new generation of farmers, technicians, and scientists, who, through evidence generated by data, make informed decisions to build profitable and sustainable agricultural systems.

## Seguridad Ciudadana y otras aplicaciones

### Test de Hipótesis para Detectar Equivalencia Estadística entre listas de Genes

Ing. Pablo Flores

## Biografía

Pablo Javier Flores Muñoz, Ingeniero en Estadística Informática por la Escuela Superior Politécnica de Chimborazo, máster en Estadística e Investigación Operativa por la Universidad Politécnica de Cataluña y Universidad de Barcelona. Es estudiante del programa de doctorado en Bioinformática por la Universidad Politécnica de Cataluña. Actualmente es coordinador de la carrera de Estadística en la ESPOCH. Trabaja en proyectos dentro de los grupos de investigación en Bioestadística y Bioinformática (Universidad de Barcelona) y grupo de investigación en ciencia de datos (ESPOCH). Ha publicado para varias revistas con indexación regional e internacional SJR. Ha participado como ponente en varios congresos nacionales e internacionales y ha sido capacitador externo para organismos y universidades como la Escuela Politécnica del Ejército, Sociedad Ecuatoriana de Estadística, entre otras.

## Resumen

Las tecnologías ómicas han revolucionado la biología y la medicina del siglo XXI. A través de los denominados “experimentos de alto rendimiento (por ejemplo, microarrays)” es posible identificar y estudiar una gran cantidad de datos de biología molecular como listas de genes, proteínas, etc.

El presente trabajo propone un test de hipótesis estadístico, que permite determinar similitud biológica entre dos listas de genes usando el enfoque de equivalencia. Justificamos la disimilitud de Sorensen como un estadístico adecuado, posteriormente, investigamos la distribución muestral teórica y aproximada de este estadístico, luego construimos un R-Package que nos permite aplicar toda la teoría investigada, el cual finalmente es usado para determinar similitudes biológicas entre listas de genes procedentes de dos experimentos reales.

## Análisis de correspondencias múltiples para el estudio de los homicidios intencionales en el Ecuador

Mauricio Abril, MSc

## Biografía

- Ingeniero Matemático por la Escuela Politécnica Nacional
- Magister en Estadística Aplicada por la Escuela Politécnica Nacional
- Estadístico Observatorio Metropolitano de Seguridad Ciudadana de Quito
- Estadístico Ministerio del Interior Plan Nacional de Seguridad
- Asesor en Estadística Fiscalía General del Estado
- Coordinador de Operaciones Empresa de Pasajeros de Quito
- Docente Universitario.
- Consultor en Minería de Datos y Estadística

## Resumen

El presente estudio, permite analizar las relaciones entre las categorías de variables asociadas con los homicidios intencionales<sup>1</sup> en el Ecuador, por medio del Análisis de Correspondencias Múltiples, para determinar las relaciones entre las modalidades que influyen y contribuyen al cometimiento de este tipo de violencia, permitiendo optimizar la elaboración e implementación de política pública, para reducir y erradicar los niveles de violencia que se tiene en el Ecuador.

## A hybrid boosting-based machine for rare events in cross-sectional studies

Prof. Dr. Montserrat Guillén

Prof. Dr. Manuela Alcañiz

Prof. Dr. Jessica Pesantez-Narvaez

## Biografía

Jessica Pesantez-Narvaez is Visiting professor at the University Carlos III de Madrid in the Statistics Department. She has received her PhD at the University of Barcelona in 2021. Since 2018 she has been member of the Riskcenter-IREA and teaching faculty in the Statistics program (UB-UPC). Her main areas of interest are artificial intelligence, theoretical and applied econometrics, and data mining.

Montserrat Guillen is Chair professor of Econometrics at the University of Barcelona and director of Riskcenter-IREA. She is Visiting Professor in the Faculty of Actuarial Science and Insurance at Cass Business School, City, University of London. She was awarded by the ICREA Academia distinction in 2011 and 2019.

Manuela Alcañiz is Associate professor at the University of Barcelona and member of Riskcenter-IREA. She has focused her research career on applied statistical methodology. She is associate editor of the Spanish Journal of Statistics.

## Resumen

A hybrid boosting-based machine learning algorithm is presented for the modelling of rare events in binary responses. It aims to (i) approximate an econometric model that allows interpretability, and (ii) reduce the prediction error of the rare phenomenon. We propose a calibration mechanism to oversample or undersample instances according to their misclassification likelihood. We also adapt a generalized least squares bias correction strategy to decrease the prediction error. The proposed method proves to improve the rate of detection of rare events tested on a real French insurance data set.

## Posters

### Ciencias Sociales

#### **Prediction of confirmed cases of and deaths caused by COVID-19 in Chile through time series techniques: A comparative study**

1, 2 Claudia Barría-Sandoval , 3 Guillermo Ferreira

1 Nursing School, Universidad de las Américas, Concepción, Chile. cbarria@udla.cl2

2 Faculty of Nursing, Universidad de Concepción, Chile. claudiabarria@udec.cl

3 Department of Statistics, Universidad de Concepción, Chile. gferreira@udec.cl

#### **Resumen**

To understand the behavior of the spread of COVID-19 as well as the projection of infections and deaths. This information is very relevant so that public health organizations can distribute financial resources efficiently and take appropriate containment measures.

#### **Referencias**

- [1] Brockwell, Peter J and Davis. Time series: theory and methods. Springer Science & Business Media; 1991
- [2] Durbin, James and Koopman, Siem Jan. Time series analysis by state space methods. Oxford university press; 2012.
- [3] Robert, Christian and Casella, George. Monte Carlo statistical methods. Springer Science & Business Media; 2013.
- [4] Dunsmuir, William TM. Generalized linear autoregressive moving average models. Handbook of discrete-valued time series. 2015: 51--76.
- [5] Holt, Charles C. Forecasting seasonals and trends by exponentially weighted moving averages. International journal of forecasting. 2004;20(1).

#### **Identificación de clusters espaciales de empresas y la influencia de factores externos en su constitución utilizando técnicas de minería de datos**

Jorge Jácome

Science Faculty, Escuela Politécnica Nacional, Quito-Ecuador, [jorge.jacome@epn.edu.ec](mailto:jorge.jacome@epn.edu.ec)

#### **Resumen**

For the Ecuadorian state as well as for the decentralized autonomous governments, it is important to have knowledge of the distribution in space of the different sectors, taking as an example the companies within the economic sector in Quito Metropolitan District.

## Referencias

- [1] Anselin, Luc. (2001). A Companion to Theoretical Econometrics. Chapter Fourteen Spatial Econometrics. Blackwell Publishing Ltd, 2001.
- [2] Celemin, Juan P. (2011). AutocorrelacionEspacial e Indicadores Locales de AsociacionEspacial. Importancia, Estructura y Aplicacion.
- [3] Goodchild, Michael F. (1986). Spatial Autocorrelation.

## Preferencias por redistribución después del auge redistributivo en América Latina

1 Pamela Caiza, 2 Yasmín Salazar

1 Facultad de Ciencias, Escuela Politécnica Nacional, Quito-Ecuador, raiza.caiza@epn.edu.ec

2 Departamento de Economía Cuantitativa, Escuela Politécnica Nacional, Quito-Ecuador, yasmin.salazar@epn.edu.ec

## Resumen

Las preferencias por redistribución se han estudiado desde diversas aristas, sin embargo aún quedan preguntas por responder. América Latina vivió un periodo de bonanza económica lo que dio lugar a la “época de oro social” en la región, pero las favorables condiciones económicas se revirtieron a finales de 2015. Entonces, en este punto es preciso preguntarse: ¿cómo se formaron las preferencias por redistribución en América Latina luego un periodo de auge redistributivo que coincidió con el inicio de un nuevo ciclo político?

## Referencias

- [1] Meltzer, A., & Richard, S. (1981). A Rational Theory of the Size of Government. Journal of Political Economy, 5, 914–927.
- [2] Salazar, Y., & Waltenberg, F. (2018). Desigualdade de Renda e Demanda por Redistribuição Caminham Juntas na América Latina no Período 1997-2015. Dados, 61(4), 341–384.
- [3] Uribe, M. (2018). Ciclos políticos y política social en América Latina en el siglo XXI. Revista Forum, (13), 101–118.

## Ciberseguridad

### Security analysis in smart home based on the OWASP ASVS methodology on a real case study

1 Roberto Andrade, 2 Jairo Chancusig, 3 Jefferson Guasumba, 4 Sang Yoo

1 Faculty of Systems Engineering, EPN, Quito-Ecuador, roberto.andrade@epn.edu.ec

2 Faculty of Systems Engineering, EPN, Quito-Ecuador, jairo.chancusig@epn.edu.ec

3 Faculty of Systems Engineering, EPN, Quito-Ecuador, jefferson.guasumba@epn.edu.ec

4 Faculty of Systems Engineering, EPN, Quito-Ecuador, sang.yoo@epn.edu.ec

## Resumen

Smart Homes area relatively new development, whose objective is to incorporate intelligence into homes to provide different types of services such as medical care, security, etc. The users of this technology are exposed to possible attacks, due to the vulnerabilities present in the devices that are part of the Smart Home.

## Referencias

- [1]Riquebourg,V.,Menga,D.,Durand,D.,Marthic,B.,Delahoche,L.,Loge,C.The Smart Home Concept: our immediate future. IEEE International Conference on E-Learning in Industrial Electronics. Available from: <https://www.infona.pl/resource/bwmeta1.element.ieee-art-000004152762/tab/summary>. (2006).
- [2] Huang,J.,Kong,L.,Chen,G.,Wu,M-Y.,Liu,X.,Zeng,P.(2019). Towards Secure Industrial IoT: Block chain System with Credit-Based Consensus Mechanism. IEEE Transactions on Industrial Informatics.
- [3]Ruef,A.,Hicks,M.,Parker,J.,Levin,D.,Mazurek,M.,Mardziel,P.(2016).Buildit, breakit, fixit: Contesting secure development. ACM SIGSAC Conference on Computer and Communications Security.

## Evaluación del riesgo de una red de IoT en una casa inteligente utilizando redes bayesianas

1 Diego Heredia, 2 Miguel Flores, 3 Roberto Andrade

1 Departamento de Matemática, Facultad de Ciencias: Escuela Politécnica Nacional, Quito-Ecuador, [diego.heredia01@epn.edu.ec](mailto:diego.heredia01@epn.edu.ec)

2 Grupo MODES, SIGTI, GEAA, Departamento de Matemática: Escuela Politécnica Nacional, Quito-Ecuador, [miguel.flores@epn.edu.ec](mailto:miguel.flores@epn.edu.ec)

3 Departamento de Informática y Ciencias de la Computación: Escuela Politécnica Nacional, Quito-Ecuador, [roberto.andrade@epn.edu.ec](mailto:roberto.andrade@epn.edu.ec)

## Resumen

La evaluación de los riesgos en las redes del Internet de las Cosas (IoT) es una tarea importante en la actualidad debido a la evolución de los ciber ataques. Las casas inteligentes, al usar redes de IoT, también están expuestas a estos ataques. Se han diseñado metodologías basadas en redes bayesianas para la seguridad de redes informáticas, sin embargo, debido a la falta de información, los parámetros no consideran las probabilidades adecuadas ni la relación entre los dispositivos de la red. Por esta razón es necesario implementar entornos de simulación que tomen en cuenta estos aspectos y junto con las redes bayesianas realizar la evaluación del riesgo.

## Referencias

- [1] Koller, D. y Friedman, N. Probabilistic Graphical Models: Principles and Techniques. Adaptive Computation and Machine Learning series. MIT Press, 2009.
- [2] Darwiche, A. Modeling and Reasoning with Bayesian Networks. Cambridge University Press, 2009.
- [3] Liu, Y. y Man, H. Network vulnerability assessment using Bayesian networks. International Society for Optics and Photonics, SPIE, 2005.
- [4] Ibrahim, M. y Nabulsi, I. Security analysis of smart home systems applying attack graph. IEEE, 2021.
- [5] NIST. Common vulnerability scoring system version 3.1 specification document. <https://www.first.org/cvss/specification-document>

## Economía y Finanzas

### Modelamiento por componentes principales con el enfoque del análisis de datos funcionales para predecir el riesgo país del Ecuador.

1 Alexander Constante, 2 Miguel Flores, 3 Paul Carrillo-Maldonado

1 Departamento de Matemática, Facultad de Ciencias, Escuela Politécnica Nacional. Quito, Ecuador, [alexander.constante@epn.edu.ec](mailto:alexander.constante@epn.edu.ec)

2 Grupo MODES, SIGTI, Departamento de Matemática, Facultad de Ciencias, Escuela Politécnica Nacional. Quito, Ecuador, [miguel.flores@epn.edu.ec](mailto:miguel.flores@epn.edu.ec)

3 Universidad de Las Américas (UDLA) y Ecuadorian Political Economy Lab (EPEL). Quito, Ecuador, [paul.carrillo.maldonado@udla.edu.ec](mailto:paul.carrillo.maldonado@udla.edu.ec)

## Resumen

Country risks determined by the Emerging Markets Bond Index (EMBI), in basis points, and is usually related as the thermometer of the health of a country's economy, with which the lower the scores investors will have more confidence to invest in the country. However, despite being a very important indicator, there are few models that predict future trends and in a multivariate context, good results have not been reflected. Therefore, the Functional Data Analysis (FDA) technique is applied to forecast future scores in a normal and extreme scenario, considering the historical information of Ecuador.

## Referencias

- [1] Banco Central del Ecuador. (2021). Indicadores Económicos. Riesgo País, diaria: <https://www.bce.fin.ec/index.php/informacioneconomica>
- [2] Hyndman, R. J. y Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. Computational Statistics & Data Analysis, 51(10):4942–4956



[3] Kokoszka, P. y Reimherr, M. Introduction to Functional Data Analysis. Chapman and Hall/CRC, Septiembre 2017.

[4] Lapitz, R., Gorfinkiel, D., Acosta, A., Flórez, M. y Gudynas, E. (2005). El Otro Riesgo País: Indicadores y desarrollo en la economía global. Uruguay: Coscoroba ediciones.

[5] Ramsay, J. y Silverman, B. Functional Data Analysis. Springer, Estados Unidos, 2005

## **Medición de la incertidumbre del Ecuador a partir del uso de técnicas de machine learning.**

1 Gabriel Amagua, 2 Miguel Flores Sánchez, 3 Paul Carrillo-Maldonado

1 Departamento de Matemáticas, Escuela Politécnica Nacional, Quito-Ecuador, gs.amagua@gmail.com

2 Departamento de Matemáticas, Escuela Politécnica Nacional, Quito-Ecuador, miguel.flores@epn.edu.ec

3 Universidad de Las Américas (UDLA) y Ecuadorian Political Economy Lab(EPEL), Quito-Ecuador, paul.carrillo.maldonado@udla.edu.ec

### **Resumen**

Es frecuente notar como se habla de cambios en los niveles de incertidumbre; sin embargo, muchas de estas opiniones se basan en impresiones y no en datos que las respalden, La falta de una medida o aproximación de la incertidumbre no permite que se pueda estudiar su trayectoria, así como tampoco estimar su impacto en el desempeño de una economía

### **Referencias**

[1] Baker, S., Bloom, N., & Davis, S. (2016). Measuring Economic Policy Uncertainty. The Quarterly Journal of Economics, 1593-1636.

[2] Padilla, Sebastián. (2019). Índice de incertidumbre de política económica para-Ecuador: Discusiones y una propuesta de cuantificación. En: Revista Puce. Issn: 2528-8156. Num.108. PP. 117-221.

[3] Echevarría Icaza, Víctor. (2019). Desentrañando las causas de la incertidumbre de política económica en España: una aproximación usando Machine Learning

## **Personalidad del emprendedor y su relación con la innovación en micro y pequeñas empresas del contexto fronterizo Ecuador - Colombia**

1 Verónica Janneth García Ibarra, 1 Luis Homero Viveros Almeida, 2 Claudia Solarte, 2 Gloria Rivera

1 Carrera de Administración de Empresas, Universidad Politécnica Estatal del Carchi, Tulcán-Ecuador, veronica.garcia@upec.edu.ec; luis.viveros@upec.edu.ec

2 Programa de Administración de Empresas, Universidad CESMAG, Pasto-Colombia, cmsolarte@unicesmag.edu.co; garivera@unicesmag.edu.co

## Resumen

The economic consequences caused by the Covid-19 pandemic have generated high rates of business closure, which is demonstrated according to data from the registry of the Single Registry of Tax payers of the Internal Revenue Service (SRI), and the Chamber of Commerce of Pasto. This implies an aggravating problem for the economy of the province of Carchi and the department of Nariño

## Referencias

- [1] Costa, P., & McCrae, R. (1992). Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO FFI): Professional Manual. Psychological Assessment Resources. <https://doi.org/10.4135/9781849200479.n>
- [2] Hambrick, D. C. (2007). Upper echelons theory: An update. *Academy of Management Review*, 32(2), 334-343. <https://doi.org/10.5465/amr.2007.24345254>
- [3] Hambrick, D. C., & Mason, P. A. (1984). Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review*, 9(2), 193–206. <https://doi.org/10.5465/AMR.1984.4277628>

## Medio Ambiente

### Homogenización de series meteorológicas mediante el uso del paquete R Climatol

1 Jonathan Pallasco

1 Facultad de Ciencias, Escuela Politécnica Nacional, Quito-Ecuador,  
jonathan.pallasco@epn.edu.ec

## Resumen

Errors in taking measurements are inherent in all activities, and meteorology is no exception. For this reason, being able to detect and correct these errors and inhomogeneities in the climatological series is very old as in the climatology itself. In ancient times, the climatological study of the series of meteorological measurements were carried out manually. Nowadays there are statistical packages with which it is possible to make the construction of complete climatological series, from an estimate of the missing or missing data, these packages work with extensive amounts of data, however, they are focused on the homogenization of series monthly, mainly temperature and precipitation, but there has been a growing interest in the homogenization of daily series, necessary for the study of the variability of extreme phenomena. The R Climatol package is one of the most up-to-date and its methodology is simple. Everything is carried out with R code and graphically shows all the processing that is being given to the information, and contains functions to homogenize, to carry out quality control and to fill in missing data in a set of series of any climatic variable.

## Referencias

- [1] Alexandersson, H. (1986). A test of homogeneity applied to the precipitation data. *Journal of Climatology*, 6(6), 661-675.
- [2] Gaona, G., Quentin, E., & Labus, J. (2013). Homogeneity and spatial variability of meteorological series of the Project area "City of Knowledge-Yachay". *ACI Advances in Sciences and Engineering*, 5(2)
- [3] Guijarro, J. (2015). CLIMATOL: FREE SOFTWARE FOR THE CLEANING AND HOMOGENIZATION OF CLIMATOLOGICAL DATA.
- [4] Toret, A., Kuglitsch, F. G., Xoplaki, E., Della-Marta, P. M., Aguilar, E., Prohom, M., and Luterbacher, J. (2011). A note on using the standard normal homogeneity test to detect inhomogeneities in climatic time series. *International Journal of Climatology*, 31(4), 630-632

## Nueva Metodología de Detección de Anomalías para Datos Funcionales

1 Jorge Sosa

1 Facultad de Ciencias, Escuela Politécnica Nacional, Quito-Ecuador, jorge.sosa01@epn.edu.ec

## Resumen

La identificación de datos atípicos es una tarea de minería de datos y más en concreto de técnicas basadas en el conocimiento descubierto en bases de datos (Knowledge Discovery in Databases, KDD). Trabajar con estos datos pueden producir distorsiones a la hora de presentar resultados, es así como, poder identificarlos y hasta saber cómo se produjeron pueden ayudar a evitar malos funcionamientos. Cuando un conjunto de observaciones discretas puede ser vista como una función, pueden tratarse mediante el Análisis de datos Funcionales (FDA), rama estadística que toma como observaciones a las funciones directamente. EIFDA es un análisis relativamente nuevo, por tanto, brindar nuevas metodologías ayudará en su progreso.

## Referencias

- [1] Papadimitriou, S., Kitagawa, H., Gibbons, P. B., & Faloutsos, C. (2003, March). Loci: Fast outlier detection using the local correlation integral. In *Proceedings 19th international conference on data engineering* (Cat. No. 03CH37405) (pp. 315-326). IEEE.
- [2] Febrero-Bande, M., Galeano, P., y González-Manteiga, W. (2007). Outlier detection in functional data by depth measures, with application to identify abnormal NO<sub>x</sub> levels. *Environmetrics*, 19(4):331-345.

## Aplicación de índices de capacidad funcional univariados para medir la productividad del maíz ante el cambio climático por medio de las variables tiempos térmicos, radiación solar y precipitación

1 Juan Játiva, 2 Miguel Flores

1 Departamento de Matemática, Escuela Politécnica Nacional, Quito-Ecuador, [juan.jativa@epn.edu.ec](mailto:juan.jativa@epn.edu.ec)

2 Departamento de Matemática, Escuela Politécnica Nacional, Quito-Ecuador, [miguel.flores@epn.edu.ec](mailto:miguel.flores@epn.edu.ec)

### Resumen

In Ecuador, climate change has been evident from 1969 to 2006 with an average increase of one degree Celsius ( $1^{\circ}\text{C}$ ), an 8% increase in precipitation in the Sierra region and 33% in the Coastal region. Impacts on agriculture have also been identified, as evidenced by the delay in the beginning and end of the rain season and heavy rain fall in short periods in the coastal region, affecting planting and harvesting dates

### Referencias

- [1] Guevara, R., Vargas, J. (2015). Process capability analysis for nonlinear profiles using depth functions. *Quality and Reliability Engineering International*, 31,465-487.
- [2] López-Pintado S, Romo J. (2007). Depth-based inference for functional data. *Computational Statistics & Data Analysis*, 51,4957-4968.
- [3] Montgomery, D. (1996). *Introduction to Statistical Quality Control*. 3rd edition. New York.: John Wiley & Sons.
- [4] Ramsay, J., y Silverman, B. (2005). *Functional data analysis*. 2nd edition. New York: Springer.

## La detección de datos anómalos de series temporales meteorológicas y climáticas mediante la profundidad funcional

1 Julián Rodríguez, 2 Miguel Flores

1 Facultad de Ciencias, Escuela Politécnica Nacional, Quito-Ecuador, [julian.rodriguez@epn.edu.ec](mailto:julian.rodriguez@epn.edu.ec)

2 Departamento de Matemática, Escuela Politécnica Nacional, Quito-Ecuador, [miguel.flores@epn.edu.ec](mailto:miguel.flores@epn.edu.ec)

### Resumen

The detection of anomalous data is a very important problem and it is essential to detect these observations, to eliminate the more to treat them. Currently there are different methods for detecting outliers. However, due to current technology, the development of Big Data and its exponential growth in recent decades, its treatment requires new methods, such as functional

data analysis (FDA) to try to reduce costs and have a better study of observations. Meteorological data measured in the Sierra region of Ecuador will be treated, as a variable relative humidity, from 2015 to 2017 per day, verifying if the geography and topology of the sector, generates atypical data and identifying the variability that exists between the stations, by applying modal depth and Bootstrap smoothing algorithm.

## Referencias

- [1]. Ramsay, J. y Silverman, B. *Functional Data Analysis*. Springer, Estados Unidos, 1997.
- [2]. Febrero-Bande, M. y de la fuente, M. O. *Statistical computing in functional*. Journal of Statistical Software, 2012.
- [3]. Hawkins, D. *Identification of Outliers*. Springer, Switzerland, 1980.

## Identification of Oil Spills on Satellite Images

1 Viviana Gavilanes Guerrero, 2 Adrián Enríquez, 3 Miguel Flores

- 1 Department of Mathematics, Escuela Politécnica Nacional, Quito-Ecuador, adrian.enriquez@epn.edu.ec
- 2 Department of Mathematics, Escuela Politécnica Nacional, Quito-Ecuador, kathya.gavilanes@epn.edu.ec
- 3 Department of Mathematics, Escuela Politécnica Nacional, Quito-Ecuador, miguel.flores@epn.edu.ec

## Resumen

Environmental protection is of public concern. As population growth, pollution and environmental accidents have increased, it is important to monitor these incidents and with special attention to oil spills and their effects. Several articles point out that early detection of oil spills is the key to making decisions to reduce their inevitable impact. Satellites such as Sentinel-1, 2 or 3 are very effective tools for analyzing the Earth's surface. These systems are powered by a Synthetic Aperture Sensor (SAR) from which the images they capture are named and have a high capacity to adapt to a variety of weather conditions. An oil spill incident that happened on August 10, 2017 in southern Kuwait is being investigated

## Referencias

- [1] BBhandari. "Processing Sentinel-1 SAR images using Sentinel Application Platform (SNAP)". En: *Geomatics, remote* (2019).
- [2] Méndez, E. "Change Detection with Synthetic Aperture Radar Imagery". En: Universidad de Zurich. Departamento de Geografía (2018).
- [3] Frank H. Wonglan G. Cumming. *Digital Processing of Synthetic Aperture Radar Data: Algorithms and Implementation*. Artech House, 2005.
- [4] Henri Maitre. *Processing of Synthetic Aperture Radar. 1.aed. SAR Images Digital Signal Image Processing Series Iste-DSP*. Wiley-ISTE.

- [5] Vermeer Martiny Tesfaye Getachew. Digital Aerial Mapping. A Hands-OnCourse, 2021.
- [6] R Nathan. "Synthetic Aperture Radar Tool and Libraries: A framework for Geo-Referenced Data Prices singand Algorithm Protoyping". En: Utah State University(2012). [7] Serco Italia SPA. "Oils pill mapping with Sentinel-1 (version1.2)". En: RUS Lectures(2019).
- [8] RWest. "Model-Based Stripmap Synthetic Aperture Radar Processing". En: Ph.D. dissertation. Utah State University (2011)

## **Predicción de la precipitación a partir de variables meteorológicas utilizando modelos de regresión funcional**

1 Danilo Loza, 2 Omar Llambo, 3 Miguel Flores

1 Department of Mathematics: Escuela Politécnica Nacional, Quito-Ecuador, danilo.loza@epn.edu.ec

2 Department of Mathematics: Escuela Politécnica Nacional, angel.llambo@epn.edu.ec

3 MODES Grup, SIGTI, GEAA, Department of Mathematics: Escuela Politécnica Nacional, Quito-Ecuador, miguel.flores@epn.edu.ec

### **Resumen**

Understand the relationship that exists between water requirements and development levels in the different stages of corn plantations. For which it is required to predict the precipitation considering that it is the natural water source of the crops.

### **Referencias**

- 1] Estrada, R. D. (2011). Ajustes al índice de potencialidad agrícola de turcpara lograr mejores diseños de los mecanismos para compartir beneficios en los andes. Technicalreport, RIMISP.
- [2] Ramsay, J. O. & Silverman, B. W. (2005).Functional Data Analysis, volume 40.Springer, New York

## **Modelización de la dinámica temporales de poblaciones de insectos**

1 Jason Paul Obando Romero

1 Facultad de Ciencias, Escuela Politécnica Nacional, Quito-Ecuador, jason.obando@epn.edu.ec

### **Resumen**

Insects perform essential functions for human life such as pollination, energy transmission or biological control. Recent publications suggest a decrease in insect abundance incertain part soft he world[1]. The cause of there ported decreases is uncertain[2].

## Referencias

- [1] Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O. U., Swartz, B., Quental, T. B., Marshall, C., McGuire, J. L., Lindsey, E. L., Maguire, K. C., Mersey, B., & Ferrer, E. A. (2011). Has the Earth's sixth mass extinctions ready arrived? *Nature*, 471(7336), 51–57. <https://doi.org/10.1038/nature09678>
- [2] Leather, S. R. (2018). “Ecological Armageddon”-more evidence for the drastic decline in insect numbers. *Ann Appl Biol*, 172, 1–3. <https://doi.org/10.1111/aab.12410>

## Herramientas de R-Shiny para el análisis estadístico interactivo temporal y espacial de variables meteorológicas monitoreadas por el GEAA (2014-2019)

1 Franklin Paucar Chicaiza, 2 Olger Capito Armas, 3 Arquímedes Haro, 4 Isabel Escudero

1, 2 Carrera de Estadística, Facultad de Ciencias, Escuela Superior Politécnica de Chimborazo, Riobamba-Ecuador, [franklins.paucar@esPOCH.edu.ec](mailto:franklins.paucar@esPOCH.edu.ec)

3 Carrera de Física, GEAA, Facultad de Ciencias, Escuela Superior Politécnica de Chimborazo, Riobamba-Ecuador,

4 Carrera de Estadística, GEAA, Facultad de Ciencias, Escuela Superior Politécnica de Chimborazo, Riobamba-Ecuador,

## Resumen

La sociedad en general tiene interés en conocer el comportamiento de las variables meteorológicas en tiempo real, es por ello que el GEAA enfatiza la aplicación de herramientas de R-Shiny para proporcionar información de forma interactiva y sencilla, mediante una interfaz gráfica y resumida de las condiciones meteorológicas de Chimborazo.

## Referencias

- [1] Berndt, C., & Haberlandt, U. (2018). Spatial interpolation of climate variables in Northern Germany—Influence of temporal resolution and network density. *Journal of Hydrology: Regional Studies*, 15, 184-202. <https://doi.org/10.1016/j.ejrh.2018.02.002>
- [2] García, F. F. (2012). Meteorología y climatología. Aspectos generales. *Índice: Revista de Estadística y Sociedad*, 50, 6-9.
- [3] Hengl, T. (2007). A practical guide to geostatistical mapping of environmental variables. Publications Office. <https://publications.jrc.ec.europa.eu/repository/handle/JRC38153> [
- [4] Vandeput, N. (2021, julio 30). Forecast KPI: RMSE, MAE, MAPE & Bias. Medium. <https://towardsdatascience.com/forecast-kpi-rmse-mae-mape-bias-cdc5703d242d>

## Negocios e Industria

### Generating vocabulary of emergencies in Spanish yielded in posts of Twitter

1 Joel García, 2 Jesús Zambrano, 3 Jorge Párraga

1 FCI, UTM, Portoviejo-Ecuador, jgarcia5169@utm.edu.ec

2 FCI, UTM, Portoviejo-Ecuador, jzambrano1217@utm.edu.ec

3 FCI, UTM, Portoviejo-Ecuador, jorge.parraga@usach.cl

#### Resumen

Extracting emergency data from Twitter is becoming a really difficult tasks in cethis is not a formal plat form to extract data and users can publish what ever they want such as irrelevant content orfakenews. Twitter has a limited number of API calls within a periodof 15 minutes, so, it's necessary to be able to extract relevant data in order to maximize the gaining of information before the limit call is reached

#### Referencias

- [1] LennCun, Y. A tutorial on Energy-Based Learning, The Courant Institute of Mathematical Sciences, New York University, 2006, extracted from paper2.dvi (lecun.com)
- [2] GM Harshvardhan, M. UBMTR: Unsupervised Boltzmann machine-based time-aware recommendation system, Journal of King Saud University -Computer and Information Sciences, 2021, extracted from<https://doi.org/10.1016/j.jksuci.2021.01.017>.

### Segmentación de Anunciantes usando Técnicas de Minería de Datos

1 Mario Aldeán

1 Departamento de Ciencias, Escuela Politécnica Nacional, Quito-Ecuador, mario.aldean@epn.edu.ec

#### Resumen

Written media such as magazines sell advertisements deeply related to the subjects their articles are about but most of the time these subjects are hand picked, based on editors hunches or novelties which might not bring reliable profit to the publisher.

#### Referencias

- [1] K. P. Sinaga and M. Yang, "Unsupervised K-Means Clustering Algorithm," in IEEE Access, vol. 8, pp. 80716-80727, 2020, doi: 10.1109/ACCESS.2020.2988796.



## Seguridad Ciudadana y Otras Aplicaciones

### Afectación de las pruebas de comparación de medias ante alteraciones de la Homocedasticidad

1 Ángeles Tipanquiza, 1 Pablo Flores

1 Escuela Superior Politécnica de Chimborazo (ESPOCH), Grupo de Investigación en Ciencia de Datos CITED, Riobamba-Ecuador

#### Resumen

Teóricamente, se espera que las pruebas de hipótesis trabajen con una Probabilidad de Error Tipo I TIEP admisible (igual al nivel de significancia) y una Potencia alta (cercana a 1), sin embargo, estudios han demostrado que, bajo ciertas condiciones, este escenario ideal no siempre es cierto. Para el caso de pruebas de comparación de dos medias, suponiendo muestras normales, nos preguntamos si la Student funciona adecuadamente cuando existe perfecta homocedasticidad, pero aún más, si el test de Welch es eficiente cuando se altera este supuesto

#### Referencias

- [1] Altman, D. G. and Bland, J. M. (1995). Statistics notes: Absence of evidence is not evidence of absence. *Bmj*, 311(7003):485. 4
- [2] Hsu, P. (1938). Contribution to the theory of “student’s” t-test as applied to the problem of two samples. *Statistical Research Memoirs*. 2
- [3] Levene, H. et al. (1960). Robust tests for equality of variances. *Contributions to probability and statistics*, 1:278–292. 7 Overall, J. E., Atlas, R. S., and Gibson, J. M. (1995). Tests that are robust against variance heterogeneity in k2 designs with unequal cell frequencies. *Psychological reports*, 76(3):1011–1017. 2
- [4] Rasch, D., Kubinger, K. D., and Moder, K. (2011). The two-sample t test: pre-testing its assumptions does not pay off. *Statistical papers*, 52(1):219–231. 3, 14, 23
- [5] Welch, B. (1951). On the comparison of several mean values: an alternative approach. *Biometrika*, 38(3/4):330–336. 23
- [6] Welch, B. L. (1947). The generalization of student’s’ problem when several different population variances are involved. *Biometrika*, 34(1/2):28–35. 1, 9
- Wellek, S. (2010). *Testing statistical hypotheses of equivalence and noninferiority*. CRC Press. 4, 8
- [7] Zimmerman, D. W. (2004). A note on preliminary tests of equality of variances. *British Journal of Mathematical and Statistical Psychology*, 57(1):173–181. 3

## Afectación sobre las pruebas de comparación de medias ante alteraciones de la Normalidad

1 Luis Chávez, 1 Pablo Flores

1 Escuela Superior Politécnica de Chimborazo (ESPOCH), Grupo de Investigación en Ciencia de Datos CITED, Riobamba-Ecuador

### Resumen

Teóricamente, se espera que las pruebas de hipótesis trabajen con una Probabilidad de Error Tipo I TIEP admisible (igual al nivel de significancia) y una Potencia alta (cercana a 1), sin embargo, estudios han de mostrado que, bajo ciertas condiciones, este escenario ideal no siempre es cierto. Para el caso de pruebas de comparación de dos medias, suponiendo homocedasticidad, nos preguntamos si *lat-Student* funciona adecuadamente cuando existe normalidad, pero aún más, si el test de Wilcoxon es eficiente cuando se altera este supuesto.

### Referencias

- [1] Altman, D. G. and Bland, J. M. (1995). Statistics notes: Absence of evidence is not evidence of absence. *Bmj*, 311(7003):485. 4
- [2] Bendayan, R., Arnau, J., Blanca, M. J., and Bono, R. (2013). Comparison of the procedures of fleishmanand ramberget al. for generating non-normal data in simulation studies. *Analesde Psicología/Annals of Psychology*, 30(1):364–371.
- [3] Blanca, M. J., Arnau, J., López-Montiel, D., Bono, R., and Bendayan, R. (2013). Skewness and kurtosis in real data samples. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 9(2):78.
- [4] Box, G. E. (1979). Robustness in the strategy of scientific model building. *Robustness in statistics*, 1:201–236.
- [5] Cochran, W. G. (1942). The 2 correction for continuity. *Iowa State College Journal of Science*, 16(1):421–436.
- [6] Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, 43(4):521–532.
- [7] Hsu, P. (1938). Contribution to the theory of “student’s” t-test as applied to the problem of two samples. *Statistical Research Memoirs*. 2
- [8] Rasch, D., Kubinger, K. D., and Moder, K. (2011). The two-sample t test: pre-testing its assumptions does not pay off. *Statistical papers*, 52(1):219–231. 3, 14, 23
- 9 Zimmerman, D. W. (2004). A note on preliminary tests of equality of variances. *British Journal of Mathematical and Statistical Psychology*, 57(1):173–181. 3

## Educación

### Nuevas técnicas basadas en Web Scraping y Text Mining para la identificación de tendencias de demanda laboral para el desarrollo de propuestas curriculares de las Instituciones de Educación Superior

1 Belén Cabezas, 2 Andrés Vinueza, 3 Lisette Zambrano

1 Facultad Latinoamericana de Ciencias Sociales, Flacso, Quito-Ecuador, ana.cabezas90@gmail.com

2 Escuela Politécnica Nacional, Quito-Ecuador, andresarmand@hotmail.com

3 Universidad Central del Ecuador, Quito-Ecuador, ltzambrano@uce.edu.ec

### Resumen

En un mundo globalizado y cambiante es necesario contar con nuevas técnicas para conocer las tendencias de demanda laboral con la aplicación de ciencia de datos, que permitan a las Instituciones de Educación Superior identificar las necesidades de este mercado de una manera dinámica para que las propuestas curriculares sean construidas o actualizadas con base a las particularidades y necesidades del mercado laboral identificadas. Actualmente, las propuestas curriculares no responden totalmente a las necesidades del mercado laboral, ya que la identificación de oferta académica, competencias, conocimientos y habilidades que requieren las instituciones del sector productivo se las realiza de una manera que no permite observar el dinamismo y exigencias actuales de este mercado. Así mismo, la estimación de profesionales requeridos para cubrir puestos de trabajo se la realiza con base a la percepción de los representantes de las empresas y no considerando la demanda real.

Esta necesidad se ha ido acrecentando con el pasar del tiempo especialmente con el despertar de la industria 4.0. y su influencia en las actividades realizadas por los sectores que componen el entramado productivo, por lo que se requiere contar con nuevas técnicas que permitan realizar monitoreo y análisis de la evolución del mercado laboral, considerando, las nuevas tendencias de reclutamiento de personal empleadas actualmente y que son realizadas por medio de plataformas digitales. Este problema puede ser resuelto con la aplicación de técnicas de Web Scraping y Text Mining para obtener y combinar información con el objetivo de identificar coincidencias de texto mediante el uso de patrones y palabras claves a fin de presentarla de una manera amigable a los usuarios que facilite la toma de decisiones.

### Referencias

[1] CES. (2020). Reglamento de Régimen Académico. [http://gaceta.ces.gob.ec/inicio.html?id\\_documento=238012](http://gaceta.ces.gob.ec/inicio.html?id_documento=238012)

- [2] Hernández Sampieri, R. (2014). Metodología de la Investigación. <http://observatorio.epacartagena.gov.co/wp-content/uploads/2017/08/metodologia-de-la-investigacion-sexta-edicion.compressed.pdf>
- [3] Izquierdo, C. (2011). Determinación del mercado objetivo y la demanda insatisfecha, cuando no se dispone de estadísticas. *Retos*, 1(1), 41-52.
- [4] Nolberto, V., & Ponce, M. (2008). Estadística Inferencial Aplicada. [https://edgarmartinlarosa.files.wordpress.com/2013/07/est\\_inf\\_aplicada.pdf](https://edgarmartinlarosa.files.wordpress.com/2013/07/est_inf_aplicada.pdf)
- [5] SENESCYT. (2019). Estadísticas de Educación Superior, Ciencia, Tecnología e Innovación. Retrieved agosto 18, 2021, from <https://siau.senescyt.gob.ec/estadisticas-de-educacion-superior-ciencia-tecnologia-e-innovacion/>
- [6] Amnislabs. (2018). Técnicas de Web scraping. <https://www.amnislabs.com/web-scraping>
- [7] Negrín, A. M. (2019). Las redes sociales como fuentes alternativas de datos geográficos: aplicación a la distribución territorial de los flujos turísticos en La Gomera. <https://riull.ull.es/xmlui/bitstream/handle/915/16652/Las%20redes%20sociales%20como%20fuentes%20alternativas%20de%20datos%20geograficos%20aplicacion%20a%20la%20distribucion%20territorial%20de%20los%20flujos%20turisticos%20en%20La%20Gomera%20.pdf?sequence=1>

## Cronograma

### Lunes 24

#### Negocios e Industria

9:00 - 9:30	Inauguración	
9:30 – 10:30	Tema: Analítica urbana	Nombre expositor: Ángeles Scetta MSc.

### Martes 25

#### Ciberseguridad

9:00 – 10:00	Tema: Seguridad Cognitiva	Nombre del expositor: Roberto Andrade MSc.
10:00 – 10:40	Tema: Detección Automática de Noticias Falsas en Español: Sátira Política Ecuatoriana	Nombre del Expositor: Nicolás Mafla
10:40 – 11:00	Break-Publicidad	
11:00 – 11:35	Tema: Analysis and Design of A Predictive Model for Phishing Detection Based on Url and Email Corpus	Nombre del Expositor: Fernanda Albán
11:35– 12:10	Tema: A security prototype for improving home security through LoRaWAN technology	Nombre del Expositor: Ing. Edwin Ávila
12:10 – 12:45	Tema: ¿Como convertir la seguridad en el ADN empresarial? - Caso de Éxito	Nombre del Expositor: Alicia Baldeón
12:45 – 13:00	Espacios Informativos	

13:00 -15:00 Presentación de Poster - Almuerzo

#### Medio Ambiente

15:00 – 16:00	Tema: Clasificación de eventos mediante funciones de correlación de pares locales para patrones de puntos espaciales.	Nombre expositor: Francisco Javier Rodríguez Cortés Ph.D.
16:00 – 16:35	Tema: Hydrosanitary management scene with COVID-19 for the city of Quito - Ecuador, using AI, prediction algorithms and Data Science	Nombre del Expositor: Mónica Delgado Yáñez
16:35 – 17:10	Tema: Adaptación del Método de Agrupación K-Medias para Datos Funcionales Correlacionados	Nombre de los Expositores: Jeysson Chuquín Alexandra Maigua

	Espacialmente con Aplicación a Datos del Índice de Vegetación de Diferencia Normalizada de los Páramos del Ecuador.	
<b>17:10 – 17:45</b>	Tema: Agriculture in the Digital Era: Present and Challenges	Nombre del Expositor: Guillermo Marcillo
<b>17:45 – 18:00</b>	Preguntas/Cierre de Jornada	
<b>Miércoles 26</b>		
<b>Economía y Finanzas</b>		
<b>9:00 – 10:00</b>	Tema: Estimación y predicción de modelos GARCH localmente estacionarios con parámetros variables en el tiempo: un enfoque computacional	Nombre del expositor: Guillermo Ferreira Ph.D.
<b>10:00 – 10:40</b>	Tema: El efecto asimétrico del precio del petróleo en la economía ecuatoriana	Nombre del Expositor: Paul Carrillo-Maldonado
<b>10:40 – 11:00</b>	Break-Publicidad	
<b>11:00 – 11:35</b>	Tema: Soluciones analíticas en banca mediante algoritmos de ciencia de datos	Nombre del Expositor: Jorge Martín Arevalillo
<b>11:35– 12:10</b>	Tema: Credit Scoring: Aplicando Técnicas de Regresión Logística y Modelos Aditivos Generalizados para una Cartera de Crédito en una entidad financiera	Nombre del Expositor: Andrés Suquillo
<b>12:10 – 12:45</b>	Tema: Modelos analíticos para optimizar pérdidas esperadas	Nombre del Expositor: Verónica Vaca
<b>12:45 – 13:00</b>	Espacios Informativos	
<b>13:00 -15:00</b>	Presentación de Poster - Almuerzo	
<b>Ciencias Sociales</b>		
<b>15:00 – 16:00</b>	Tema: Aplicación de Modelos Geográficamente Ponderados a la Solución de Problemas Sociales Usando R	Nombre del expositor: Fabio Sepúlveda-Murillo Ph.D.
<b>16:00 – 16:45</b>	Tema: Construction of a spatial inequality social indicator by Kernel Nonlinear Principal Components	Nombre del Expositor: Jared Abigail Valencia

		Analysis, using the Survey of Living Conditions 2014	
<b>16:45 – 17:30</b>		Tema: Identificación de clusters espaciales de empresas y la influencia de factores externos en su constitución utilizando técnicas de minería de datos	Nombre del Expositor: Jorge Jácome
<b>17:30 – 18:00</b>		Preguntas/Cierre de Jornada	
<b>Jueves 27</b>			
<b>Educación</b>			
<b>9:00 – 10:00</b>		Tema: Cultura Estadística e Investigación Cuantitativa: desafíos pendientes en el contexto de la Educación Superior.	Nombre expositor: Sergio Alberto Castillo Páez, Ph.D.
<b>10:00 – 10:45</b>		Tema: Evaluación bibliométrica de los antimicrobianos y las contribuciones en WOS (2000-2021)	Nombre del Expositor: Iván Alberto Analuisa Aroca
<b>10:45 – 11:15</b>		Break-Publicidad	
<b>11:15 – 12:00</b>		Tema: Retos en la reutilización de datos para la lucha anticorrupción, un estudio descriptivo desde la Sociedad Civil	Nombre del Expositor: Susana Cadena Ph.D.
<b>12:00 – 12:45</b>		Tema: Modelo de predicción del rendimiento académico para el curso de nivelación de la Escuela Politécnica Nacional a partir de un modelo de aprendizaje supervisado	Nombre del Expositor: Karen Calva
<b>12:45 – 13:00</b>		Espacios Informativos	
<b>13:00 - 15:00</b>		Presentación de Poster - Almuerzo	
<b>Conferencias- Talleres R Users Group</b>			
<b>Negocios e Industria</b>			
<b>15:00 – 15:45</b>		Tema: Nuevas aportaciones para el análisis de datos funcionales complejos	Nombre del expositor: Manuel Oviedo Ph.D.
<b>15:45-16:30</b>		Taller: Análisis de datos funcionales (FDA) en R utilizando el paquete fda.usc	Nombre del expositor: Manuel Oviedo Ph.D.
<b>16:30 – 17:15</b>		Tema: La Comunidad de R como generadora de sinergias entre Academia e Industria	Nombre expositor: Emilio López Cano Ph.D.

<b>17:15- 18:00</b>	Taller: Creación de blogs estadísticos con el paquete {blogdown}	Nombre expositor: Emilio López Cano Ph.D.
<b>18:00 – 18:30</b>	Presentación Colegio de Científicos de Datos y R Users Group Ecuador	

### Viernes 28

#### Seguridad Ciudadana y otras aplicaciones

<b>9:00-9:45</b>	Tema: Test de Hipótesis para Detectar Equivalencia Estadística entre listas de Genes	Nombre del Expositor: Pablo Flores
<b>9:45- 10:30</b>	Tema: A hybrid boosting-based machine for rare events in cross-sectional studies	Nombre del Expositor: Prof. Dr. Jessica Pesantez-Narvaez
<b>10:30- 11:15</b>	Tema: Análisis de correspondencias múltiples para el estudio de los homicidios intencionales en el Ecuador	Nombre del Expositor: Mauricio Abril
<b>11:15- 11:30</b>	Break-Publicidad	
<b>11:30- 12:30</b>	Tema: Modelos estadísticos espacio-temporales para el análisis, predicción y seguimiento de datos sobre delitos.	Nombre expositor: Jorge Mateu Ph.D.
<b>12:30- 13:00</b>	Clausura	