

Máster en Data Science

Propuesta de TFM

**“MODELO DE MACHINE LEARNING Y GIS PARA DETERMINAR RECURSOS CERCANOS EN LA REPUESTA DE EMERGENCIA”**

**Estudiante**: CARLOS AUGUSTO CARREÑO VILLARREYES

**15 Octubre, 2020**

**Datos Generales**

|  |  |  |  |
| --- | --- | --- | --- |
| **Documento** | **Versión** | **Autor** | **Control de Cambios** |
| UAH-TFM-Carlos-Carreno-Version-1.0.0.docx | 1.0.0 | Carlos Carreño | original |
| UAH-TFM-Carlos-Carreno-Version-1.0.1.docx | 1.0.1 | Carlos Carreño | modificado |
|  |  |  |  |
|  |  |  |  |

## Indice

[Indice 3](#_Toc49699367)

[Datos Identificativos 4](#_Toc49699368)

[Breve Introducción 5](#_Toc49699369)

[Objetivos 6](#_Toc49699370)

[Alcance de los datos y las técnicas a utilizar 6](#_Toc49699371)

[Autoevaluación del grado de innovación 9](#_Toc49699372)

[Referencias 10](#_Toc49699373)

## Datos Identificativos

A continuación, los datos identificativos de la propuesta de fin de master:

|  |
| --- |
| **Titulo** |
| MODELO DE MACHINE LEARNING Y GIS PARA DETERMINAR RECURSOS CERCANOS EN LA REPUESTA DE EMERGENCIA |
| **Estudiante** |
| Carlos Augusto Carreño Villarreyes |
| **Fecha de Propuesta** |
| 2020-09-01 |

## Breve Introducción

Gestionar la emergencia es muy crítico, en algunos casos puede resultar en una cuestión de sobrevivencia; su importancia es tal que se han creado a nivel de gobierno entidades y organizaciones para gestionar eventos de emergencia a nivel nacional. Los eventos de emergencia son aquellos que ponen en peligro el funcionamiento normal de los servicios de las empresas, organizaciones públicas, ponen en riesgo los recursos, la vida y la salud de los ciudadanos, en general son todos los eventos que ponen en riesgo la estabilidad del estado de un sistema [1].

Las personas y organizaciones que gestionan la emergencia usan herramientas para administrar los recursos antes, durante y después de un evento catastrófico. La información correcta y a tiempo, es una parte critica para el éxito de la gestión de la emergencia [3]. Los sistemas de información geográfica o GIS por sus siglas en ingles proveen la información y herramientas para el análisis de datos espaciales.

La tendencia actual en usar servicios de nube permite a las aplicaciones acceder a recursos de computo, anchos de banda y almacenamiento necesario para procesar grandes conjuntos de datos (Datasets) [2]. Las aplicaciones GIS producen y analizan grandes cantidades de Datasets que pueden contener un gran número de ficheros y objetos de datos complejos.

En la gestión de la emergencia, se utilizan sistemas de información geográfica para gestionar los recursos disponibles, en este entorno el contar con métodos precisos y algoritmos que permitan la identificación, clasificación y recuperación de los datos frecuentemente distribuidos y heterogéneos son muy críticos, tales métodos pueden provenir de la ciencia de los datos.

Nuestra propuesta de trabajo de fin de master se centra en presentar un modelo que permita la integración de la información recopilada y métodos de la ciencia de los datos para proporcionar los conocimientos necesarios para el proceso de representación visual y análisis de los eventos peligrosos, esto permite la notificación oportuna utilizando diversas interfaces del estado actual de todos los parámetros relevantes.

## Objetivos

Son objetivos de la presente propuesta:

* Identificar los métodos de la ciencia de los datos para la recopilación, procesamiento y almacenamiento de los datos de los recursos disponibles utilizados en los sistemas de gestión de emergencia.
* Proponer un modelo basado en técnicas de machine learning para determinar los recursos más cercanos a un evento de emergencia localizado para la respuesta de la emergencia.
* Crear un prototipo de validación acotado del modelo de la arquitectura propuesta

## Alcance de los datos y las técnicas a utilizar

Los datos a utilizar corresponden a los recursos disponibles para atender una emergencia a nivel nacional en la República del Perú. Los datos son administrados por la entidad INDECI (Instituto de Defensa Civil).

Los datos de los recursos disponibles están agrupados (14) en catorce categorías, llamadas también capas o temas. Los temas por los cuales se agrupan los datos son:

* Abastecimiento de agua
* Abastecimiento de Alimentos
* Abastecimiento de Energía
* Atención Medica
* Transporte y Viabilidad
* Telecomunicaciones
* Centros de decisión e intervención
* Áreas potenciales para albergues
* Áreas potenciales para escombreras
* Áreas Económicas
* Peligro sísmico
* Peligro de Tsunami
* Accesibilidad

Cada recurso puede tener una cantidad de atributos de datos diferente incluso si pertenecen a una misma categoría, así, por ejemplo: si tratamos de caracterizar una laguna usaremos diferentes atributos que para caracterizar a un rio, siendo ambos recursos de la misma capa en este caso abastecimiento de agua, incluso los datos espaciales son distintos, pues en el caso de la laguna usaremos un polígono y en el caso del rio una línea para representarlos geométricamente. Entre los atributos de los recursos se puede distinguir un conjunto de atributos comunes a todos los recursos estos serán consideradas las variables del Dataset de estudio:

|  |  |  |
| --- | --- | --- |
| Atributo | Descripción | Tipo |
| nombre | Nombre del recurso | String |
| Latitud | Latitud de la localización del recurso | Numérico |
| Longitud | Longitud de la localización del recurso | Numérico |
| Categoría | Categoría, capa o tema del recurso | String |
| Descripción | Descripción del recurso | String |
| Cantidad | Cantidad del recurso | Numérico |
| Geometría | Objeto geométrico que representa el recurso puede ser línea, polígono o punto | Espacial |
|  |  |  |

Los datos están actualmente en cerca de 200 tablas en una base de datos relacional con soporte a datos espaciales como es postgresql, la idea es explorar si hay una forma mejor de almacenar los datos por ejemplo en una base de datos NO SQL como MongoDB. El volumen de datos asciende a 5 GB para el propósito del aprendizaje automático. El conjunto de dato se ha obtenido mediante convenio con INDECI (Instituto Nacional de Defensa Civil del Peru).

Los métodos de machine learning o aprendizaje automático se pueden utilizar para mejorar la calidad de los datos y el procesamiento de los datos GIS para obtener mejores resultados [4]; así utilizaremos el algoritmo de aprendizaje supervisado de redes neuronales llamado Perceptron multicapa para automatizar la clasificación de los recursos.

En el procesamiento de las consultas a los datos identificamos que podemos utilizar el algoritmo de machine learning de aprendizaje no supervisado llamado k-medias para localizar los recursos más próximos dada una coordenada media.

## Autoevaluación del grado de innovación

|  |  |
| --- | --- |
| Aspecto | Grado de innovación |
| Datos utilizados | El Dataset de recurso es único, se compone de un conjunto de atributos singular y de geolocalización. |
| Objetivos analíticos | Los estudios existentes identifican la integración de GIS y Machine Learning, en nuestro estudio amplio el contexto para enfocarlo a la gestión de emergencia. Este enfoque resulta innovador. |
| Técnicas aplicadas para la  preparación de datos. | Las técnicas utilizadas se basan en datos relacionales, en general buscamos una alternativa para mejorar el tratamiento de los datos. En este caso buscamos una alternativa explorando las técnicas de big data y de BD NoSQL, este enfoque resulta innovador. |
| Técnicas aplicadas analíticas | Las técnicas analíticas utilizadas se basan en métodos de aprendizaje automático como el Perceptron multicapa para la clasificación. Este enfoque resulta innovador. |
| Visualización avanzada | En el campo de la visualización nos basaremos en los visores geográficos a los cuales agregaremos funcionalidades adicionales de consulta, búsqueda e interfaces de acceso al procesamiento de aprendizaje automático. |

## Referencias

1. Gamala A, Makram N. "Proposed Model of GIS-based Cloud Computing Architecture for Emergency System". 2016 Julio 25.
2. Stoimenov L, Predić B, Mihajlović V, Stanković M."GIS Interoperability Platform for Emergency Management in Local Community Environment". 2005 Enero.
3. Stoimenov L, Milosavljević A, Stanimirović A. "GIS as a Tool in Emergency Management Process". 2007 Julio 2
4. Reza S. Szekely P."Design of a GIS-based Assistant Software Agent for the Incident Commander to Coordinate Emergency Response Operations". 2014 Enero