

PROYECTO DE INVESTIGACIÓN SEMILLA PIS-17-10
"Minería de datos, extracción de vector de características y obtención de modelos con reconocimiento de patrones y aprendizaje automático para detección de escenarios relacionados con el delito de tráfico de personas"

En la ciudad de Quito D.M., a los tres días del mes de agosto del año dos mil veinte, comparecen a la celebración de la presente Acta de Finalización del Proyecto de Investigación Semilla **PIS-17-10 "Minería de datos, extracción de vector de características y obtención de modelos con reconocimiento de patrones y aprendizaje automático para detección de escenarios relacionados con el delito de tráfico de personas"**, por una parte la **Ph.D. Alexandra Patricia Alvarado Cevallos** en calidad de **Vicerrectora de Investigación, Innovación y Vinculación** de la Escuela Politécnica Nacional, y por otra la **Ph.D. Myriam Hernández** en calidad de **Directora del Proyecto de Investigación Semilla PIS-17-10**, al tenor de lo siguiente:

1. ANTECEDENTES:

- a) El 4 de julio de 2017, el Consejo de Investigación y Proyección Social mediante Resolución 079/17, aprueba el Cronograma para el lanzamiento de la Convocatoria para la presentación de Proyectos de Investigación Internos, Semilla, Junior y Multi e Interdisciplinarios 2017, y mediante Resolución 092/17 del 1 de agosto del 2017 se aprueba una reforma al Cronograma de la Convocatoria de Proyectos de Investigación del 2017; con lo cual se aplaza la fecha de cierre de la convocatoria.
- b) El 12 de diciembre de 2017, al amparo de lo dispuesto por Consejo de Investigación y Proyección Social, mediante Resolución 167/17, se aprobaron los proyectos del Informe Final - Convocatoria 2017, entre ellos el denominado "*Minería de datos, extracción de vector de características y obtención de modelos con reconocimiento de patrones y aprendizaje automático para detección de escenarios relacionados con el delito de tráfico de personas*", presentado por la Ph.D. Myriam Hernández.
- c) Mediante Memorando EPN-VIPS-2018-0553-M, del 13 de marzo de 2018, se informa a los Directores de los proyectos Semilla 2017 que la fecha de inicio de los proyectos es el 9 de abril del 2018.

2. DATOS GENERALES DEL PROYECTO:

Código de Proyecto	PIS-17-10
Nombre del Proyecto	<i>Minería de datos, extracción de vector de características y obtención de modelos con reconocimiento de patrones y aprendizaje automático para detección de escenarios relacionados con el delito de tráfico de personas</i>
Directora del Proyecto	HERNÁNDEZ ÁLVAREZ MYRIAM BEATRIZ
Codirector del Proyecto	AGUIAR PONTES JOSAFÁ DE JESÚS
Departamento	Informática y Ciencias de la Computación (DICC)
Líneas de Investigación	Machine Learning
Objetivo	Generar un corpus con información referente al delito de tráfico de personas en Ecuador, procesarlo, extraer características y clasificarlo para obtener modelos de posibles escenarios relacionados con el mencionado tipo de crimen
Duración del Proyecto	<ul style="list-style-type: none"> • Inicio: 9 de abril del 2018 • Fin planificado: 9 de octubre del 2019 • Prórroga Ordinaria: 6 meses, hasta el 8 de abril de 2020 • Duración total: 24 meses
Entrega del Informe Final	1 de junio del 2020



Presupuesto asignado	\$ 9.735,88 USD (nueve mil setecientos treintaicinco dólares americanos, con 88/100)
Presupuesto ejecutado	\$ 8.312,13 USD (ocho mil trescientos doce dólares americanos, con 13/100)

3. INFORME FINAL:

Mediante Memorando EPN-PIS-17-10-2020-0007-M del 1 de junio de 2020, la Ph.D. Myriam Hernández, Directora del Proyecto PIS-17-10, presenta el Informe Final del Proyecto Semilla, y mediante Memorando EPN-PIS-17-10-2020-0009-M del 21 de junio de 2020, presenta información complementaria. Esta documentación es revisada por la Dirección de Investigación, se anexa y forma parte integrante del Acta de Finalización, cuyas conclusiones y productos generados son:

CONCLUSIONES:

- El alcance y resultados del proyecto se mantuvieron dentro de lo planificado. Se configuró un corpus de mensajes de Twitter y de imágenes enlazadas con esos mensajes. Los mensajes se detectaron como sospechosos de promover servicios ilícitos realizados por menores de 14 años que estarían relacionados con tráfico de personas.
- El enfoque aplicado para el descubrimiento de estos mensajes tuvo dos fases: En la primera se minaron los sitios usando hashtags ya probados como relacionados con este crimen. Los tuits fueron procesados para eliminar ruido y normalizarlos. Con esta información, se extrajeron características para ingresar a un algoritmo SVM para clasificar los textos como sospechosos o no. Se construyó una lista negra de mensajes para que sus imágenes sean procesadas.
- En la segunda fase, se descargaron las imágenes incrustadas en los mensajes y las encontradas en los sitios web enlazados.
- En el problema de la clasificación por género y edad, se tiene la complejidad de que las imágenes, en muchos casos, no tienen la información de características faciales, sino solamente los datos del torso o parte superior del cuerpo, su tamaño y proporciones. Como resultado de este proyecto de investigación, se lograron excelentes resultados para detectar edad y género, aún en los casos en que no estaban disponibles las características del rostro porque se encontraban pixeladas o borradas, como sucede normalmente en esta clase de sitios.

PRODUCTOS:

- Artículo: "Detection of Possible Illicit Messages Using Natural Language Processing and Computer Vision on Twitter and Linked Websites"; Granizo S., Valdivieso A., Barona L., Hernández M.; Revista: "IEEE Access" (Indexada Scopus Q1); ISSN: 21693536; DOI: 10.1109/ACCESS.2020.2976530; marzo 2020.
- Conferencia con revisión por pares: "Gender and age classification based on human features to detect illicit activity in suspicious sites" (Indexada Scopus); Álvarez Myriam; 2019 International Conference on Computational Science and Computational Intelligence (CSCI); Las Vegas, Estados Unidos; diciembre 2019.
- Conferencia con revisión por pares: "Detection of possible human trafficking in Twitter" (Indexada Scopus); Álvarez Myriam; ICI2ST – JISIC 2019. Conferencia Internacional en Sistemas de Información y Tecnologías de Software y las X Jornadas de Ingeniería en Sistemas Informáticos y de Computación; Quito, Ecuador; noviembre 2019.

- Conferencia con revisión por pares: "*Detection of Human Trafficking Ads in Twitter Using Natural Language Processing and Image Processing*" (Indexada Scopus); Álvarez Myriam; Advances in Intelligent Systems and Computing 1213 AISC; Quito, Ecuador; julio 2020.
- Tesis de Maestría de Software: "*Procesamiento de imágenes de Twitter o URLs asociados, clasificados como sospechosos de estar relacionados con trata de personas*"; Granizo Pilataxi Sergio Leonardo.

4. LIQUIDACIÓN ECONÓMICA:

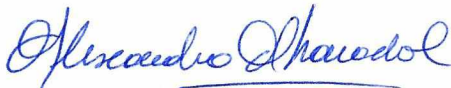
El monto asignado al Proyecto Semilla PIS-17-10 fue de \$ 9.735,88 USD (nueve mil setecientos treintaicinco dólares americanos, con 88/100), y se ejecutaron \$ 8.312,13 USD (ocho mil trescientos doce dólares americanos, con 13/100), conforme al detalle emitido por la Unidad de Gestión de Investigación y Proyección Social del Vicerrectorado de Investigación, Innovación y Vinculación, que se adjunta a la presente Acta y forma parte integrante de la misma.

5. FINALIZACIÓN:

Con la presente Acta se declara finalizado y cerrado el Proyecto Semilla PIS-17-10: "*Minería de datos, extracción de vector de características y obtención de modelos con reconocimiento de patrones y aprendizaje automático para detección de escenarios relacionados con el delito de tráfico de personas*".

Para constancia de lo ejecutado y por estar de acuerdo con el contenido de la presente Acta, las partes libre y voluntariamente suscriben la misma, en tres ejemplares de igual contenido, tenor y valor legal.

Dado en la ciudad de Quito, D.M. a los tres días del mes de agosto del año dos mil veinte.



Ph.D. Alexandra Alvarado
Vicerrectora de Investigación,
Innovación y Vinculación

sp/cr



Ph.D. Myriam Hernández
Directora del Proyecto
PIS-17-10

Recibido
21/06/2021

