RESUMO - III. PESQUISA CIENTÍFICA RELACIONADA ÀS TECNOLOGIAS EDUCACIONAIS E ASSISTENCIAIS

OTIMIZAÇÃO DA TRIAGEM POR MEIO DE MACHINE LEARNING NA ATENÇÃO PRIMÁRIA DE URGÊNCIA

Katiuska Reynaldos-Grandon (katiuska.reynaldos@unab.cl)

Daniel Aguayo Villegas (daniel.aguayo@unab.cl)

Joaquín Ureta-Achurra (j.uretaachurra@uandresbello.edu)

Erick Cortez-Pinto (erick.cortez@usach.cl)

Javier Rojas-Ávila (javier.errojas@gmail.com)

María Cecilia Opazo (mopazod@udla.cl)

Introdução: A superlotação nos Serviços de Urgência da atenção primária (SAR) é um problema crítico agravado pela ambiguidade dos registros de motivos de consulta. Esta falta de estrutura dificulta a avaliação clínica e aumenta o risco de diagnósticos errôneos em ambientes de alta pressão.

Objetivo: Otimizar a triagem por meio de Machine Learning na atenção primária de saúde de urgência.

Método: Utilizou-se um modelo de linguagem de larga escala com dados de 174.593 atendimentos realizados durante um ano em um SAR. Analisou-se a lacuna de informações entre a narrativa inicial do paciente e seu diagnóstico de alta. Foi desenvolvido um sistema analítico de três etapas: Padronização: Uso do Gemini LLM para traduzir os motivos de consulta não estruturados em códigos CID-10. Processamento: Aplicação de técnicas de Processamento de Linguagem Natural (TF-IDF) e agrupamento (K-Means) para categorizar sintomas. Validação: Visualização com t-SNE e comparação com diagnósticos médicos finais, apoiada por um aplicativo de curadoria para especialistas humanos.

Resultados: A análise de clusters identificou com sucesso quatro arquétipos distintos de pacientes que impulsionam o volume do centro: condições crônicas e complicações, doença febril aguda, traumatismo agudo e dor de curta duração, e dor orgânica não traumática. O estudo também destacou a necessidade de remover as "palavras vazias" (stopwords) administrativas, que inicialmente enviesavam a formação dos clusters.

Conclusão: A conversão sistemática dos motivos de consulta não estruturados dos pacientes em perfis clínicos processáveis fornece a base para o apoio à triagem em tempo real. A integração desses modelos facilita a geração de perguntas automáticas e específicas, otimizando a alocação de recursos e acelerando a precisão diagnóstica.

Abstract

Introduction:

Overcrowding in primary emergency care services (SAR) is a critical issue exacerbated by the ambiguity of patient chief complaint records. This lack of structure hinders clinical evaluation and increases the risk of diagnostic errors in high-pressure environments.

Objective:

To optimize triage through Machine Learning within emergency primary healthcare settings.

Method:

A large-scale data approach was used, analyzing 174,593 encounters over one year at a SAR facility. The study examined the information gap between the patient's initial narrative and their discharge diagnosis. A three-step analytical system was developed: Standardization: Use of the Gemini LLM to translate unstructured chief complaints into ICD-10 codes. Processing: Application of Natural Language Processing (NLP) techniques, specifically TF-IDF and K-Means clustering, to categorize symptoms. Validation: Visualization via t-SNE and comparison with final medical diagnoses, supported by a custom curation app for human expert review.

Results: Cluster analysis successfully identified four distinct patient archetypes driving facility volume: Chronic Conditions & Complications, Acute Febrile Illness, Acute Trauma & Short-Duration Pain, and Non-Traumatic Organic Pain. The study also highlighted the necessity of removing administrative "stopwords," which initially skewed cluster formation.

Conclusion: The systematic conversion of unstructured patient complaints into actionable clinical profiles provides the foundation for real-time triage support. Integrating these models facilitates the automatic generation of targeted questions, optimizing resource allocation and accelerating diagnostic accuracy.

Resumen

Introducción:

La saturación en los Servicios de Urgencias de atención primaria (SAR) es un problema crítico agravado por la ambigüedad de los registros de motivos de consulta. Esta falta de estructura dificulta la evaluación clínica y aumenta el riesgo de diagnósticos erróneos en entornos de alta presión.

Objetivo: Optimizar del Triage mediante Machine Learning en atención primaria de salud de urgencia.

Método: Uso de modelo de lenguaje de gran escala de datos, de 174.593 atenciones durante un año en un SAR. Se analizó la brecha informativa entre la narrativa inicial del paciente y su diagnóstico de egreso. Se desarrolló un sistema analítico de tres pasos. 1) Estandarización: Uso de Gemini LLM para traducir los motivos de consulta no estructurados a códigos CIE-10. 2) Procesamiento: Aplicación de técnicas de procesamiento de lenguaje natural (TF-IDF) y agrupamiento (K-Means) para categorizar síntomas. 3.Validación: Visualización con t-SNE y comparación con diagnósticos médicos finales, apoyada por una aplicación de curación para expertos humanos.

Resultados: El análisis de clústeres identificó con éxito cuatro arquetipos de pacientes distintos que impulsan el volumen de pacientes en el centro: afecciones crónicas y complicaciones, enfermedad febril aguda, traumatismo agudo y dolor de corta duración, y dolor orgánico no traumático. El estudio también destacó la necesidad de eliminar las palabras vacías administrativas, que inicialmente sesgaban la formación de clústeres.

Conclusión, la conversión sistemática de los motivos de consulta no estructurados de los pacientes en perfiles clínicos procesables proporciona la base para el apoyo al triage en tiempo real. La integración de estos modelos facilita la generación de preguntas automáticas y específicas, optimizando la asignación de recursos y acelerando la precisión diagnóstica.

Palavras-chave: triage; machine learning; atenção primária; urgência.