

ESTUDIO PRELIMINAR SOBRE EL USO DE machine learning (ML) PARA LA IDENTIFICACIÓN DE Bacillus spp.

CORTESE, Iliana J.; MADRASSI, Lucas M.; CASTRILLO, María L.; ZAPATA, Pedro D.

Universidad Nacional de Misiones, Facultad de Ciencias Exactas, Químicas y Naturales, Instituto de Biotecnología Misiones “Dra. María Ebe Reca” (InBioMis), Laboratorio de Biotecnología, Molecular. Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). cortesejulietta@gmail.com

INTRODUCCIÓN

La identificación de microorganismos es un requerimiento fundamental para la investigación de tecnologías que abarcan desde estudios clínicos, hasta ensayos del sector agroindustrial y ambiental. Tras el advenimiento de la era molecular, la identificación taxonómica de microorganismos incorporó el estudio de loci con relevancia filogenética. Sin embargo, los métodos tradicionales basados en Distancias Genéticas (DG) requieren de experticia, consumen elevado tiempo, recursos computacionales y no siempre producen resultados consistentes. En ese contexto, la utilización de técnicas de machine learning (ML) basadas en el uso Clasificadores (C), como el Multinomial Naive Bayes (MNB), y en técnicas de procesamiento del texto y del lenguaje natural (PLN), es una alternativa con potencial para clasificar microorganismos con base en caracteres moleculares, por ejemplo, secuencias de ADN. Estos algoritmos NO requieren de alineamientos múltiples y, una vez entrenados, pueden brindar resultados de manera rápida y accesible. Sin embargo, su implementación requiere de la búsqueda de las variables Óptimas (hiperparámetros) que maximizan la eficiencia del proceso de entrenamiento (OP).

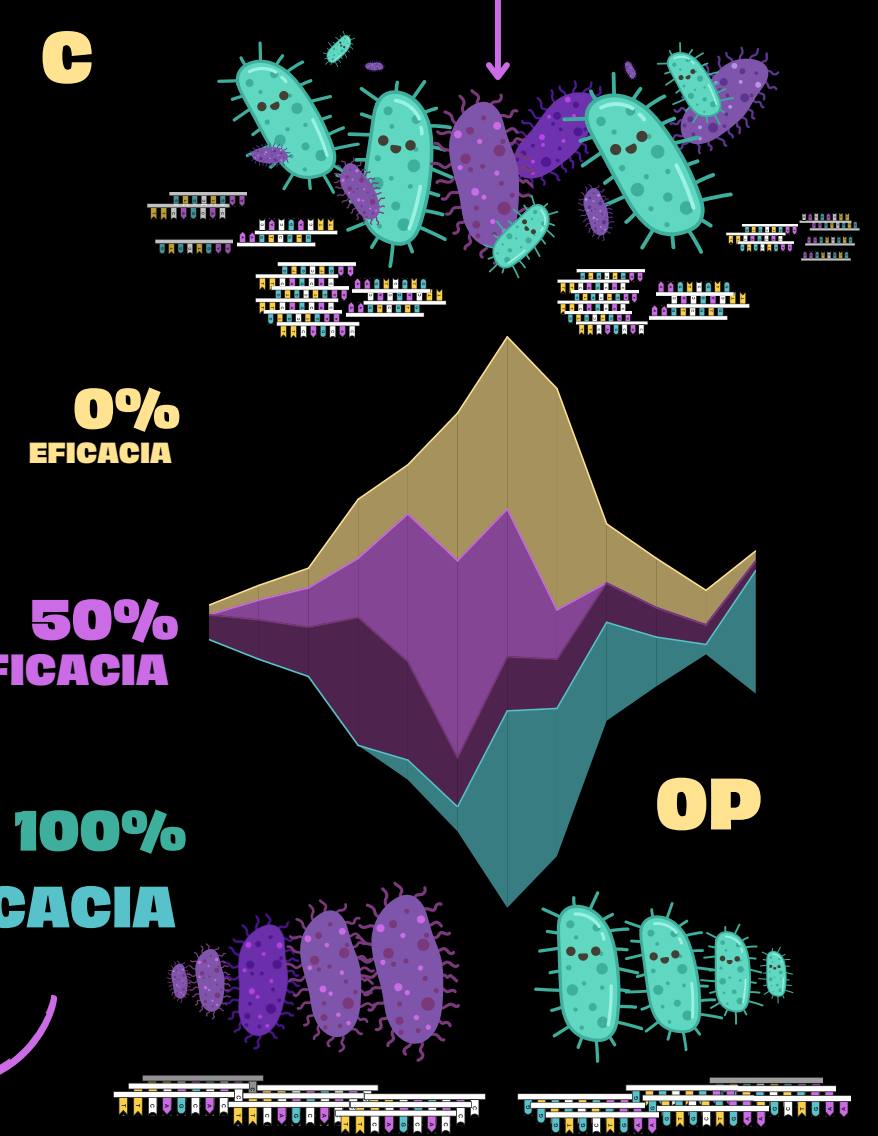
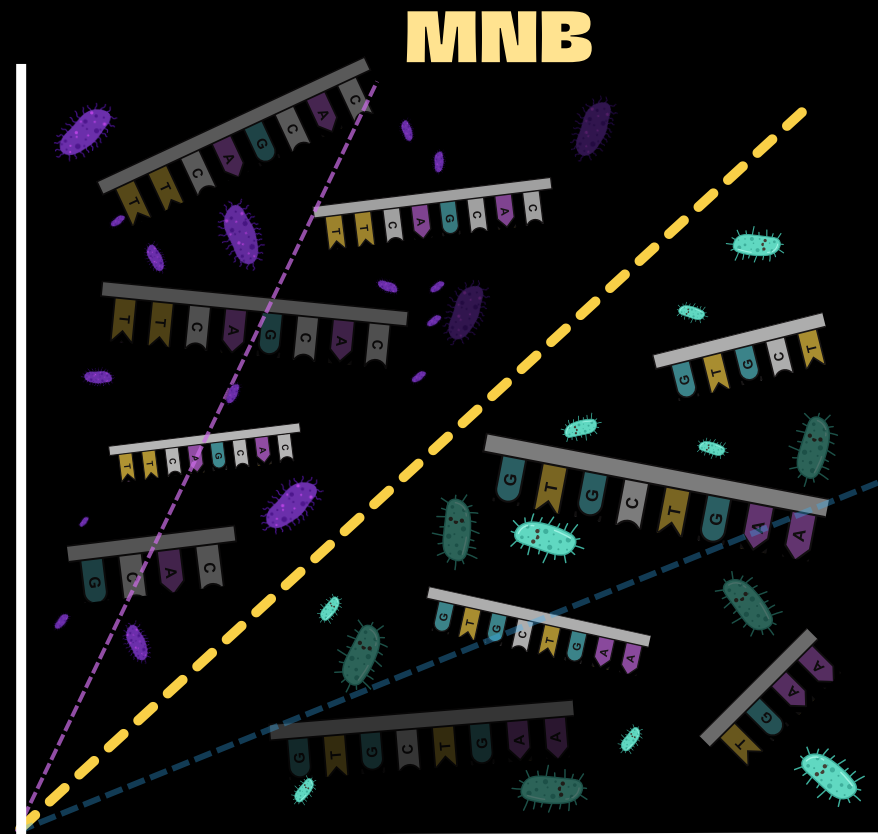
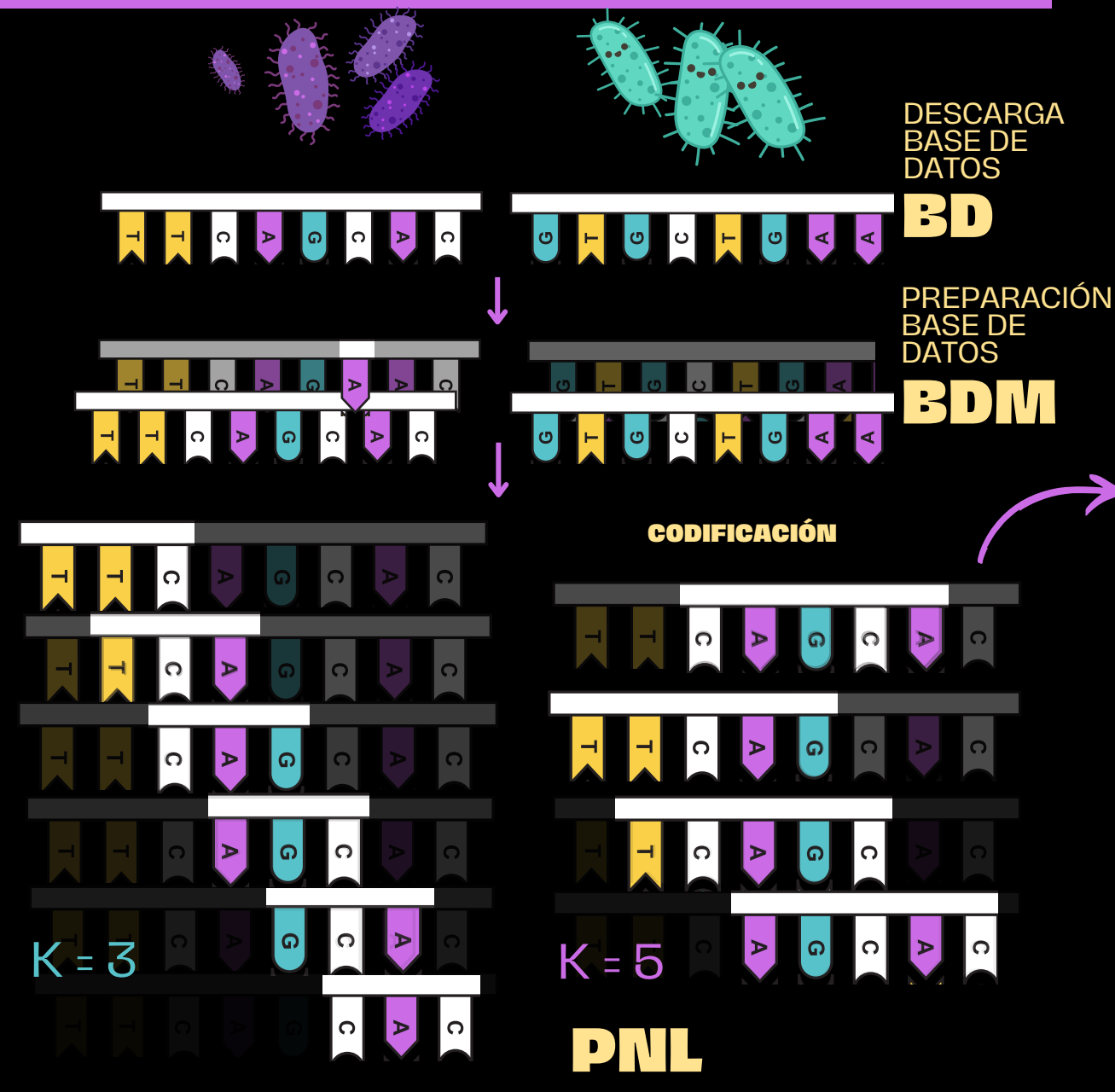
OBJETIVO

Diseñar un marco para la optimización e implementación de modelos ML mediante un clasificador MNB basados en secuencias de ADN del gen 16S del ADN ribosomal del género Bacillus.

CONCLUSIONES

El presente trabajo arroja evidencia preliminar acerca de la aplicabilidad del clasificador MNB. Sobre secuencias del gen 16S en el género Bacillus, la confiabilidad de estos métodos fue superior a de los métodos DG. la información obtenida requiere de validación en un entorno real e INTEGRACIÓN con los procesos filogenéticos basados en DG.

MATERIALES Y METODOS



RESULTADOS

Método	DG	ML	
		BD en bruto	BD editada
Base de datos (BD)	BD editada	BD en bruto	BD editada
Confiabilidad	50-80 %	60-95 %	70-100 %