Uncover New Patient Insights With Natural Language Processing at Scale

Approximately 80% of healthcare data is unstructured, and much of that data, such as provider notes, is text. Buried within are valuable insights that provide a more complete view into a patient's health. Unfortunately, most healthcare and life sciences organizations struggle to process and analyze their massive volumes of unstructured text. Databricks and John Snow Labs, the creator of the open source Spark NLP for Healthcare, help organizations transform their text data into novel patient insights with the power of natural language processing (NLP).

Innovating research and patient care with NLP

Leading healthcare and life sciences organizations use Databricks to process and analyze their text data for innovative use cases.

AstraZeneca

Enabled researchers to quickly and efficiently search thousands of research documents with machine learning and text mining

MUSC Health
Medical University of South Carolina

Analyzed over 600,000 telehealth records to identify patients at risk for COVID-19 and deploy early intervention efforts
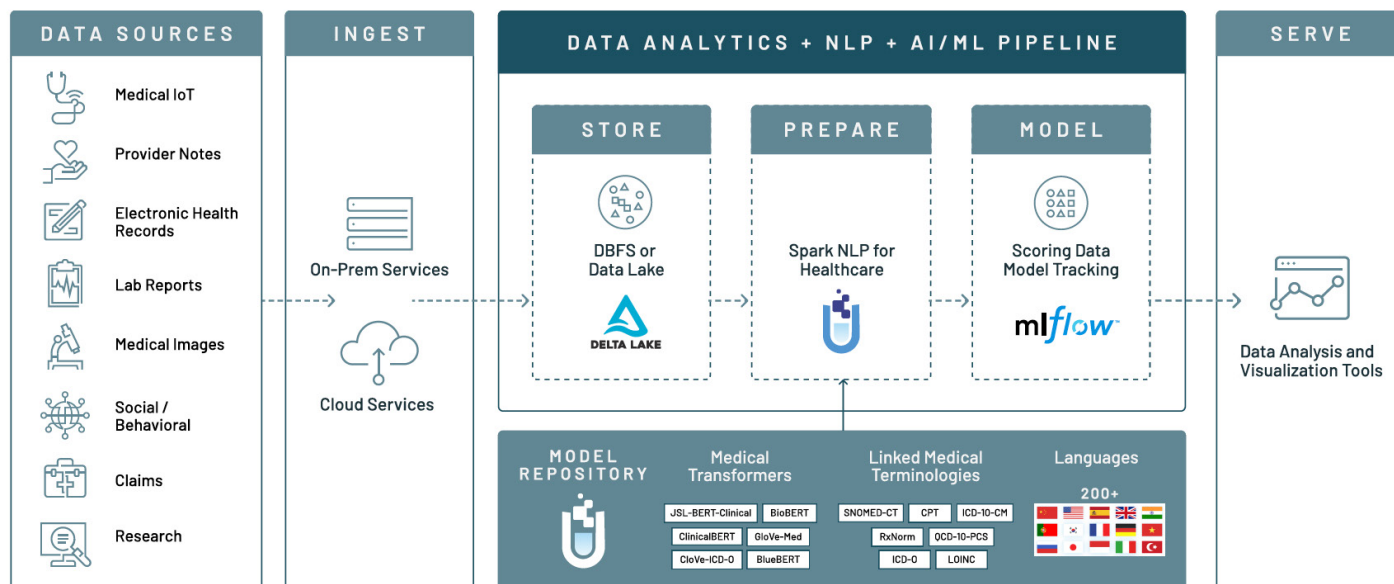
SANOFI

Analyzed claims, EHR and lab data for 17M patients to identify indications for new treatments

Transform your health text into powerful insights with Databricks and John Snow Labs

The joint solution from Databricks and John Snow Labs makes it easy to analyze text data at scale by bringing together best-of-breed healthcare NLP tools with a scalable Lakehouse Platform for all your analytics and AI. Our Solution Accelerator templates for popular NLP use cases get your data teams started faster.

databricks LAKEHOUSE PLATFORM

DATA SOURCES	INGEST	DATA ANALYTICS + NLP + AI/ML PIPELINE			SERVE
Medical IoT	On-Prem Services	**STORE**	**PREPARE**	**MODEL**	Data Analysis and Visualization Tools
Provider Notes	Cloud Services	DBFS or Data Lake — DELTA LAKE	Spark NLP for Healthcare	Scoring Data Model Tracking — mlflow	
Electronic Health Records					
Lab Reports					
Medical Images					
Social / Behavioral					
Claims					
Research					

MODEL REPOSITORY

Medical Transformers: JSL-BERT-Clinical, BioBERT, ClinicalBERT, GloVe-Med, CloVe-ICD-O, BlueBERT

Linked Medical Terminologies: SNOMED-CT, CPT, ICD-10-CM, RxNorm, QCD-10-PCS, ICD-O, LOINC

Languages: 200+

Benefits

INDUSTRY-LEADING TOOLS FOR HEALTHCARE NLP — John Snow Labs' Spark NLP for Healthcare is the most widely used natural language processing library for the healthcare and life sciences industries. It seamlessly extracts, classifies and structures clinical and biomedical text data with state-of-the-art accuracy, offering more than 200+ pretrained models.

PROCESS TEXT DATA AT MASSIVE SCALE — Built in the cloud with Apache Spark™ and Delta Lake under the hood, Databricks brings massive scale and speed to data and AI workloads. By pairing Databricks with John Snow Labs, organizations can scale their NLP workloads for billions of provider notes.

ANALYZE ALL YOUR HEALTH DATA — Databricks provides a unified platform for all your data (structured, semi-structured and unstructured), analytics and AI. Our joint solution enables healthcare organizations to easily connect their text data with all health data — such as, medical images, genomics, EHR — to create a holistic view of their patients.

NLP use cases in healthcare and life sciences

PROVIDER
- Care Delivery
- Care Pathways
- Automated Coding
- Population Health

PAYER
- Prior Authorization Review
- Billing Audits
- Medicare Risk Adjustment
- Care Management

PHARMA
- Clinical Trial Optimization
- Comparative Effectiveness
- Brand Commercialization
- Safety/Surveillance

Get started with Solution Accelerators for NLP

Our Solution Accelerators help your teams get started faster with prepackaged analytics templates that include sample data, prebuilt code, and step-by-step instructions ready to go in a Databricks notebook.

GENERATING RWE FOR ONCOLOGY
Extract critical information from PDF pathology reports and leverage it in use cases ranging from measuring disease progression to exploring optimal treatment pathways.

AUTOMATING DATA DE-IDENTIFICATION
One of the biggest challenges in healthcare is anonymizing Protected Health Information (PHI). Use this solution to automate the process of removing PHI from tabular, free-text, scanned and imaging data.

About John Snow Labs Inc.

John Snow Labs Inc., the creator of the open source Spark NLP library, Spark NLP for Healthcare and Spark OCR, is a leading healthcare AI company. John Snow Labs' Spark NLP for Healthcare is the most widely used NLP library in the healthcare industry. It seamlessly extracts, classifies and structures clinical and biomedical text data with state-of-the-art accuracy, offering more than 200+ pretrained models.

Learn more about healthcare and life sciences solutions

DBRICKS.CO/LIFESCIENCES

DBRICKS.CO/HEALTHCARE