CASE STUDY

Advancing Precision Medicine through Transformative Text Analysis in Pharmaceutical Research

CONTATA

OVERVIEW

The client is one of the world's top 5 drug manufacturing and pharmaceutical companies headquartered in Japan with a global presence in more than 80 countries.

CHALLENGES

The client partnered with Contata to automate the curation of disease-gene-mutation relationship from the vast pool of unstructured images and text. The information was either found in biomedical literatures or curated databases. Doctors require access to this information because it allows them to understand how gene-mutations are linked with various genetic diseases. Existing, human-curated databases were struggling to keep up with the rapidly evolving landscape of research findings, leading to a significant gap in providing up-to-date and comprehensive insights.

INDUSTRY

• Pharmaceutical

SERVICES

• Document Analytics
• Automation

TECHNOLOGY

• NLP
• Machine Learning

SOLUTION

Leveraging advanced Named Entity Recognition (NER) technology, Contata enabled the identification and extraction of relevant entities from biomedical literature, streamlining the identification of gene-mutations and diseases. We employed state-of-the-art machine learning (ML) algorithms to facilitate the extraction of two-way and three-way relationships between genes, mutations, and diseases, ensuring a comprehensive understanding of their interconnectedness.

Our solution also included using ML algorithms to accurately identify negations within the text, reducing the risk of misinterpretation and ensuring the precision of curated information. To enhance data consistency and comparability, Contata implemented entity normalization techniques, ensuring that entities across different sources were represented consistently.

We developed a robust mapping system to link mutation mentions to specific mutation IDs, providing a standardized approach to referencing genetic variations. Contata implemented a ranking matrix to prioritize gene variants associated with genetic diseases, facilitating efficient decision-making for healthcare professionals. Utilizing advanced text mining approaches, our solution established clear associations between genes and diseases, contributing to a more nuanced understanding of genetic factors in disease pathology.

BENEFITS

- **Timely Access to Critical Information**: Doctors and researchers now have rapid access to the latest and most relevant information on gene-mutation-disease relationships.
- **Data Accuracy and Consistency**: Automation significantly reduced the risk of human error, ensuring a high level of accuracy and consistency in the curated information.
- **Up-to-Date Research Insights**: The client was now up-to-date with the latest research findings, providing a competitive edge in biomedical research.
- **Enhanced Efficiency in Precision Medicine**: The streamlined curation process and the ranking matrix optimized the precision medicine approach.

About Contata

Contata Solutions is a trusted leader in technology and digital innovation. Through our work in data engineering, data analytics, machine learning, marketing automation and app development, we deliver solutions that address complex problems in ways that are simple, insightful and impactful.

Our promise and value proposition to our customers is simple: we leverage our deep technical expertise and global presence to bring software products and data-driven decision capabilities to life.

Founded in 2000, Contata is a privately-held company headquartered in Minneapolis that serves clients globally from offices in the United States and India.