



Data Science & Analytics

**DS/ML  
capabilities in  
E-Protocol  
module in APOGI**

[www.geninvo.com](http://www.geninvo.com)

North America | UK | Europe | India



# Business Need

---

## Auto extraction of Meta data from Protocol documents

- ▶ Meta data such as Trial design datasets, study visit schedule etc., forms the basis of CRF generation and SDTM data transformation. At present ,this meta data is being extracted manually from the protocol.
- ▶ This module aims to auto extract the meta data using NLP techniques such as NER model, Q&A model and semantic search.



[www.geninvo.com](http://www.geninvo.com)  
North America | UK | Europe | India





# Challenges



## Text Extraction

- Pdminer
- Grobid



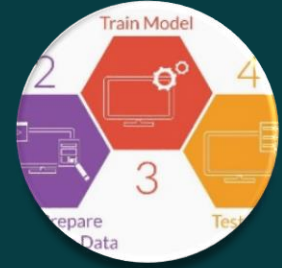
## Table Extraction

- Pdf plumber
- Tabula



## Information Retrieval

- NER(Spacy)
- Q&A(BERT)
- Semantic Search(Sentence-Transformer)



## Preparation of training data

- Web scraping- from Clinicaltrials.gov



www.geninvo.com  
North America | UK | Europe | India

# Business Impact

---

- ▶ Reduces the manual efforts for metadata extraction so that user can focus on tasks such as CRF generation and data transformation.
- ▶ Greater focus on automation which in turn helps user in long run



[www.geninvo.com](http://www.geninvo.com)  
North America | UK | Europe | India







Ideas  
Delivered  
Intuitively

**Thank You!**

---

[www.geninvo.com](http://www.geninvo.com) | Reach out to us [info@geninvo.com](mailto:info@geninvo.com)

