

Identifying Similar Insurance Claims using Text-Based Vector Search

Daniel Jakobi → Markel Insurance SE,
daniel.jakobi@markel.de

Davide De March → Markel International,
davide.demarch@markel.com

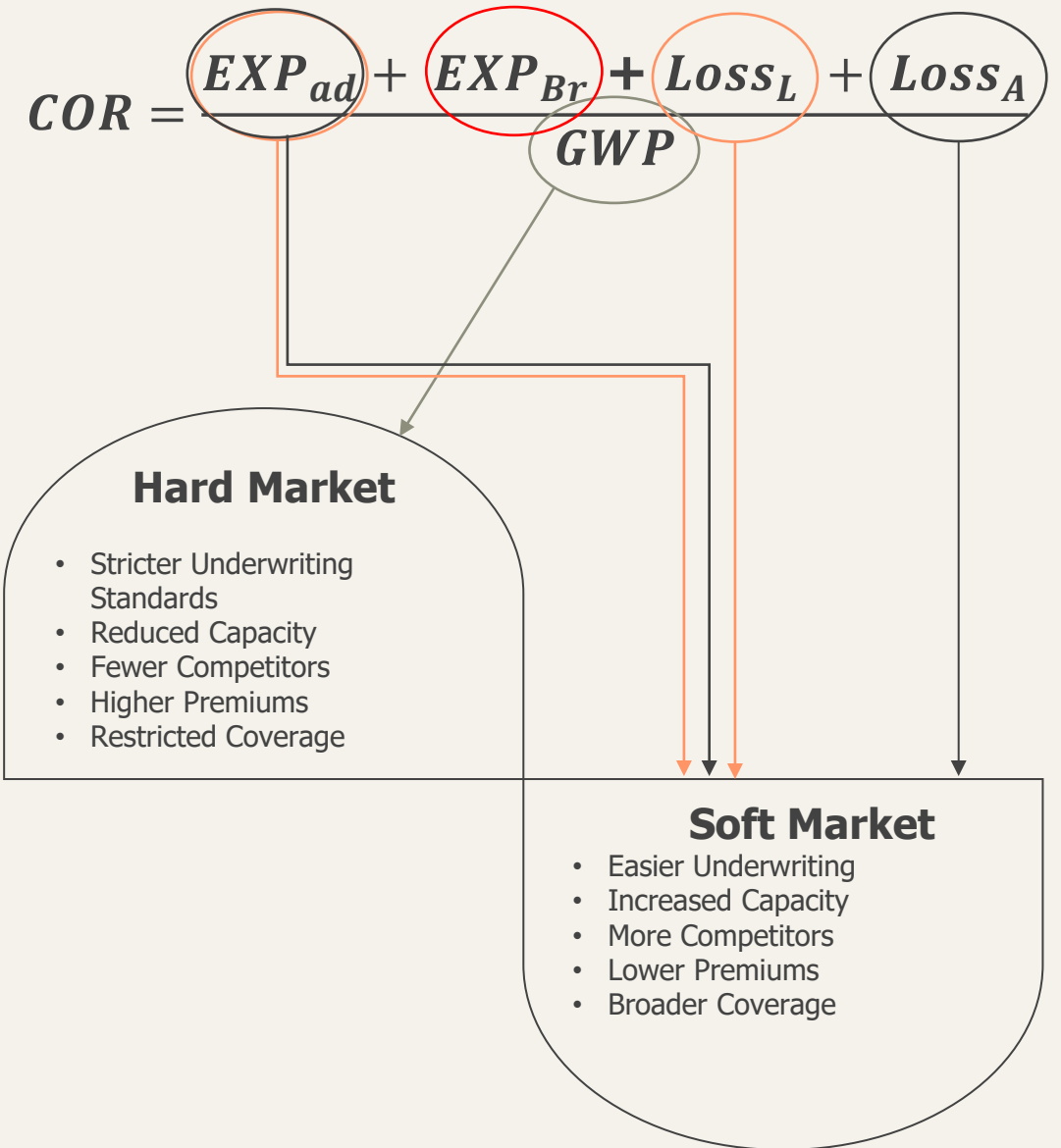
19 June 2025



Support the business with the help of AI

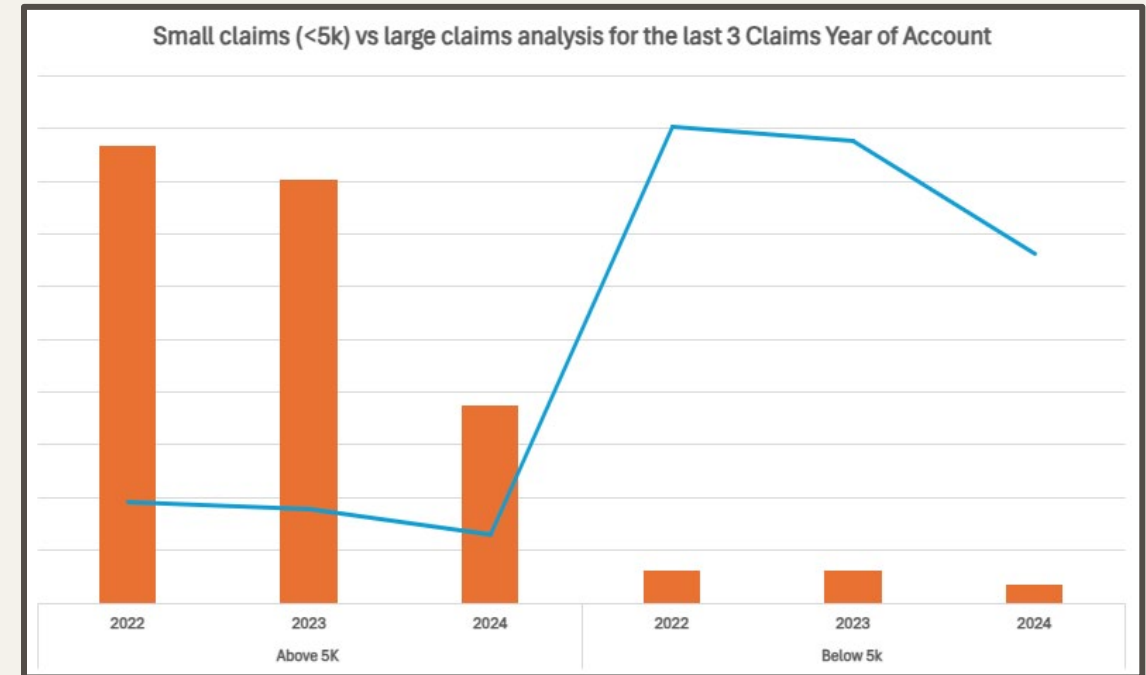


Claim AI – Why focus on Claims?

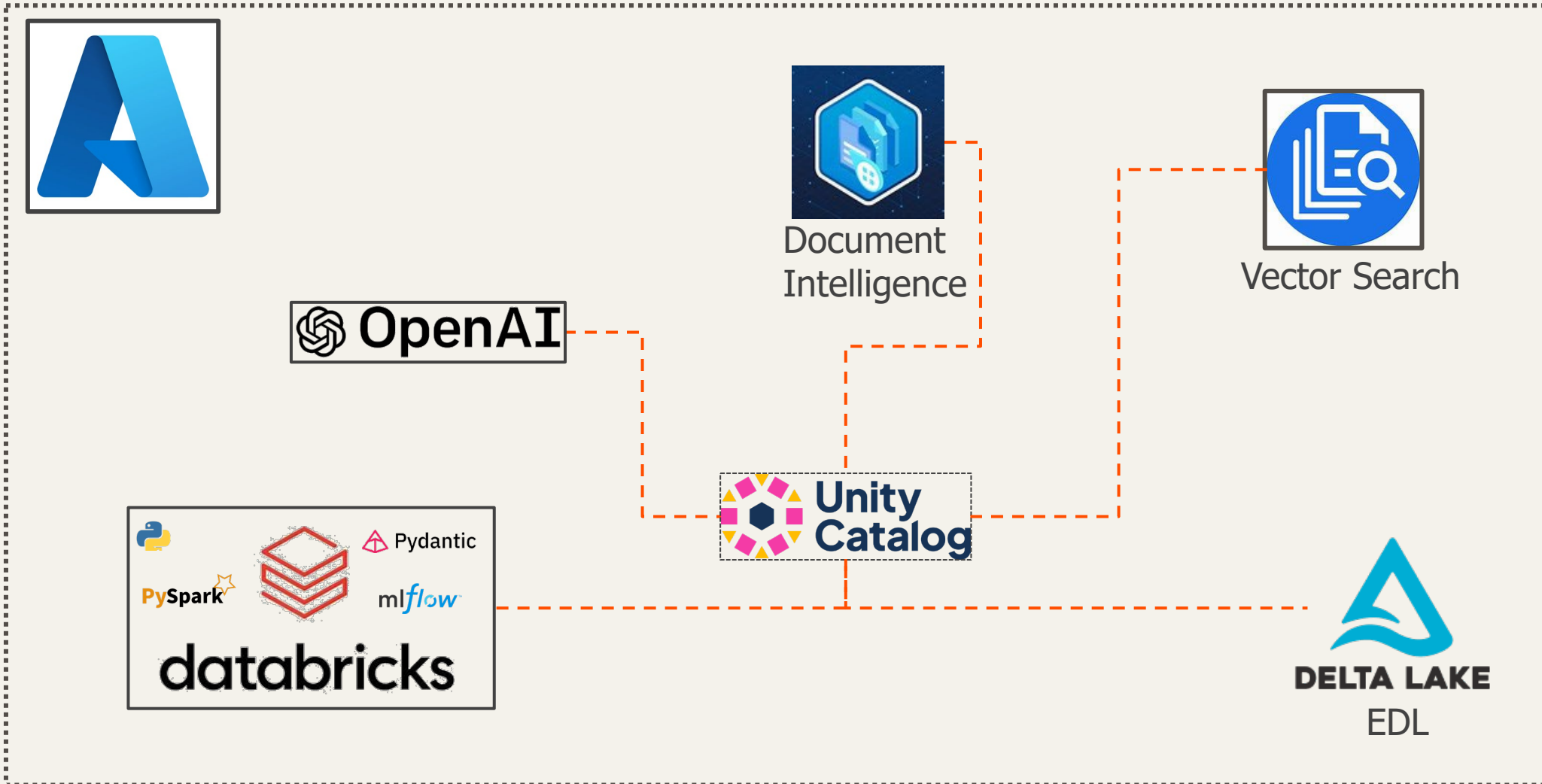


Efficiency boost and loss ratio reduction through AI: Break the link between # of claims and # of claims' handler

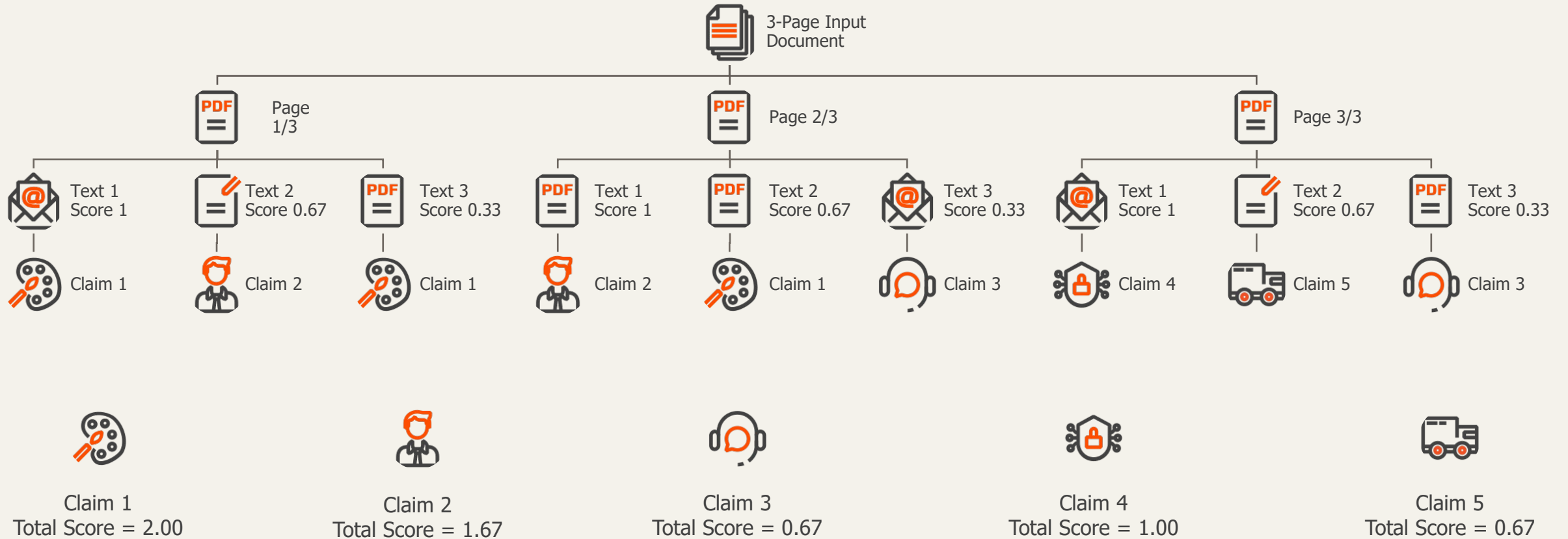
- Vast majority of claims are within 0 and 5k
- Claims handler needs to spend time to read through email and documents to assess the claims
 - While they can focus more on large claims where we end up having unexpected losses
- The earlier we identify those large claims
 - the more leverage we have in the settlement process
 - and avoid litigation
- Claims handler can spend more time on important claims and let technology help dealing with high-volume low payment claims
- AI can help
 - provide evidence on large claim,
 - highlights key sentences
 - and find claims' similarity in Market



Technical Stack



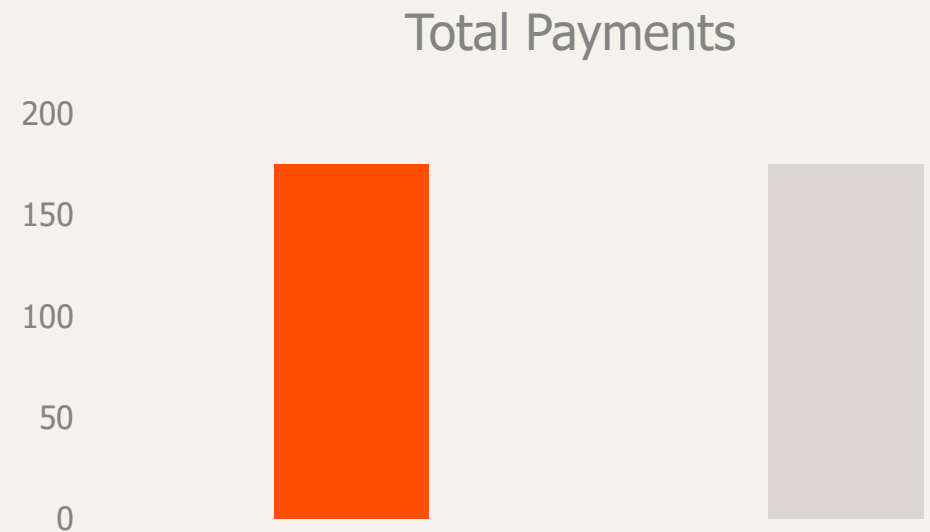
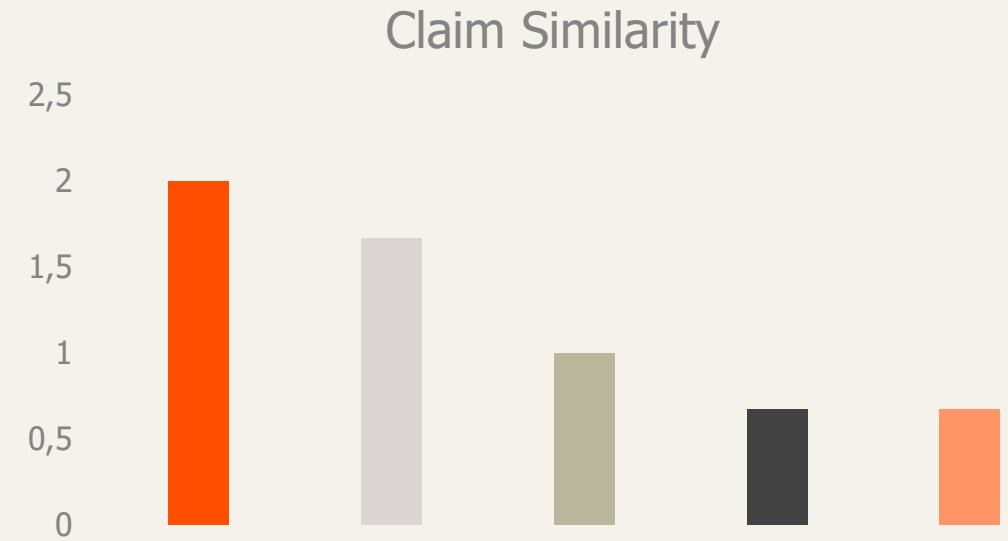
How to get from document similarity to claim similarity



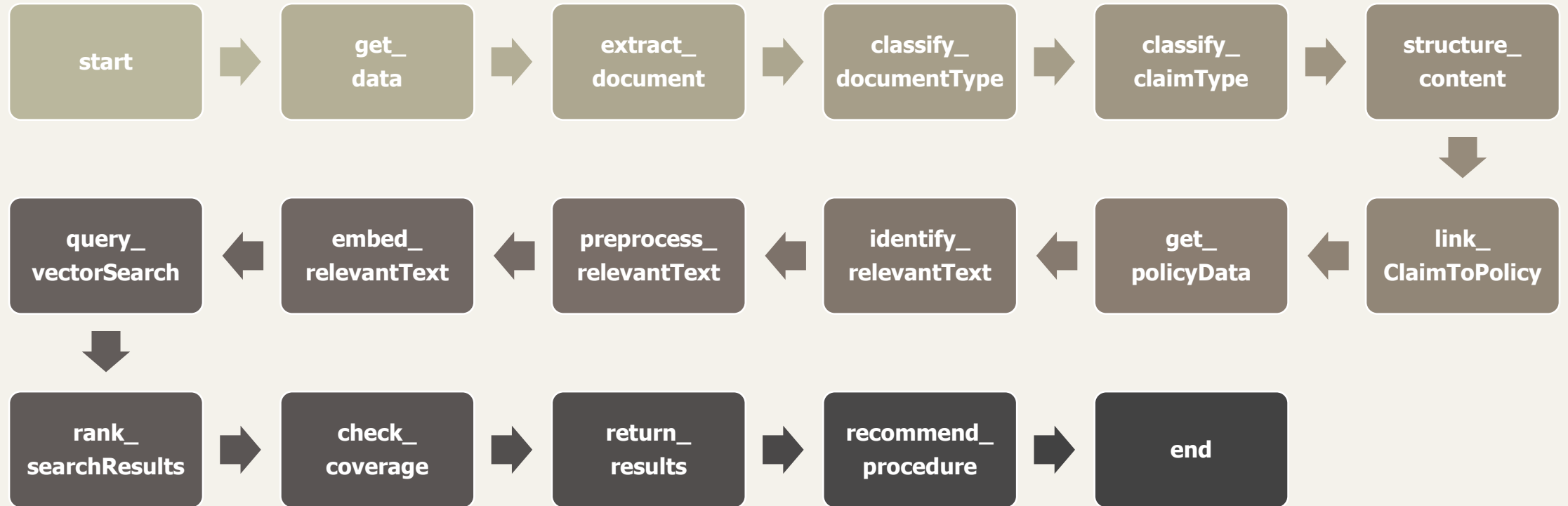
Overview of similarity and deductible

- we paid for 2/5 most similar claims
- paid amount is lower than deductible for 3/5 of similar claims
- this can potentially be automated

Policy ID	Risk	Deductible	
Pol_1	Risk_1	0	
Pol_2	Risk_1	1000	
Pol_3	Risk_1	250	
Pol_4	Risk_1	250	
Pol_5	Risk_1	500	



Claims AI – Graph structure



Impacts

Process
→

Vector Search Database

- Build a Market library of all the historic claims notes and descriptions (leveraging multilingual embeddings)
- Compare new claims with all previous similar cases based on claims notes in the library
- Predict size of the risk and help triage quickly, settle if needed, or autopay for very small claims

LLM process

- Read text, scanned and digital documents to extract a dictionary of information
- Check coverage with RAG
- Email response with claims details can be sent out in minutes
- Response to Client within 24/48 hours 100% of the time

Data
→

API Data Extraction from claim system

- Email and attachment are stored in data lake and linked with IDs
- Claims notification is logged in the system in structured way

Leverage DataLake

- New claims link immediately to Policy info (Limit/excess/deductible) in data lake and wordings
- Easy to identify exclusion

Questions?

→ daniel.jakobi@markel.de

→ davide.demarch@markel.com

MARKEL