Afstudeeropdracht: Geautomatiseerde codering van medische aansprakelijkheidsclaims

# Doel

MediRisk verzekert, zonder winstoogmerk, ziekenhuizen en andere zorgorganisaties voor het risico van medische aansprakelijkheid. MediRisk behandelt schadeclaims en werkt samen met haar leden actief aan het voorkomen van schade. De historische database, met gegevens van meer dan 25.000 claims ingediend sinds de oprichting in 1993, speelt hierbij een centrale rol. Na de indiening van elke claim, worden tientallen relevante kenmerken handmatig geabstraheerd (“gecodeerd”) uit het juridische en medische dossier en toegevoegd aan deze database door een medisch codeur. MediRisk wil onderzoeken in hoeverre deze handmatige codering (deels) kan worden geautomatiseerd met behulp van *natural language processing* (NLP) technieken.

# Achtergrond

Wanneer een behandeling in een ziekenhuis leidt tot onverwacht letsel bij een patiënt, als gevolg van een medische fout, kan een patiënt het ziekenhuis aansprakelijk stellen voor de geleden schade. Deze schade kan betrekking hebben op verlies van arbeidsvermogen (bijvoorbeeld wanneer een patiënt niet meer kan werken door blijvende invaliditeit), kosten van verzorging en huishoudelijke hulp (wanneer een patiënt door het opgelopen letsel niet langer voor zichzelf kan zorgen), of smartengeld. De geclaimde schade kan variëren van enkele honderden tot miljoenen euro’s per geval.

Ziekenhuizen verzekeren zich tegen dergelijke claims bij een verzekeraar van medische aansprakelijkheid, zoals MediRisk, en betalen hiervoor jaarlijks een bepaalde premie. Wanneer een ziekenhuis een claim ontvangt kan het deze doorsturen naar de verzekeraar, en deze laatste handelt de claim namens het ziekenhuis af en keert, bij erkenning van aansprakelijkheid, een vergoeding uit aan de patiënt. Het is ook mogelijk dat het ziekenhuis/MediRisk aansprakelijkheid afwijst, omdat er geen sprake is van een (verwijtbare) medische fout, maar bijvoorbeeld van een complicatie. Ook kan aansprakelijkheid worden afgewezen, omdat er geen verband kan worden aangetoond tussen de medische fout en het letsel. Tenslotte, kan de uitkomst van een claim ook zijn dat een minnelijke regeling wordt getroffen. In dat geval wordt geen aansprakelijkheid erkend, maar ontvangt de patiënt wel een vergoeding.

# Het coderingsproces

Ongeacht het standpunt dat MediRisk inneemt ten aanzien van de aansprakelijkheid, wordt de claim gecodeerd. Dat betekent dat een medisch codeur het claimdossier doorneemt, en een aantal relevante kenmerken van de claim invoert in de claimdatabase, zoals:

* Hoofdspecialisme (chirurgie, gynaecologie etc.)
* Afdeling (verloskunde, lab, poli etc.)
* Type zorgproces (diagnostiek, operatie etc.)
* Behandeld orgaan (rug, onderarm, hart etc.)
* Getroffen orgaan (idem)
* Gevolg
* Oorzaak (communicatie)

Bij de meeste kenmerken kan worden gekozen uit circa 10-20 verschillende waarden, waarbij in sommige gevallen meer dan één waarde kan worden gekozen (wanneer bijvoorbeeld aan meerdere organen letsel is ontstaan, of er meer dan één oorzaak is).

# Aanpak

MediRisk wil onderzoeken in hoeverre deze handmatig ingevoerde kenmerken geautomatiseerd kunnen worden gelabeld op basis van natural language processing van het onderliggende juridische en medische dossier. Dit bevat onder meer:

* De brief met aansprakelijkstelling door de claimant
* Het medische dossier van de claimant
* Eén of meerdere medische adviezen gegeven door de medisch adviseur van MediRisk omtrent de mate van aansprakelijkheid
* Communicatie over het dossier van de claimbehandelaar en betrokkenen

Bovenstaande documenten zijn beschikbaar als ascii tekst, verkregen na digitalisering, OCR en pseudonimisering van de oorspronkelijke papieren dossiers. In totaal zijn ongeveer 12000 dossiers op deze manier beschikbaar, gemiddeld bevat één dossier 200 pagina’s.

Een mogelijke aanpak zou kunnen zijn een NLP algoritme te trainen (supervised learning) op een set reeds gecodeerde claims (dwz. waarvan de kenmerken reeds zijn ingevoerd in de database, en de bovenstaande documenten digitaal beschikbaar zijn).

# Praktische toepassing

Een succesvol algoritme zou in staat moeten zijn:

* Kenmerken van toekomstige claims automatisch te labelen (bijv: getroffen orgaan = pols, oorzaak = communicatie)
* Aan te geven wat de mate van betrouwbaarheid is van elk toegekend label (zodat de medisch codeur selectief controles kan uitvoeren)