CONCEPT VAN EEN SYSTEEM VOOR PRODUCTRELATIE-ANALYSE

Samuel Jendryke

XENOS Schutweg 5, 5145 NP Waalwijk

Luty, 2026

Inleiding

Het doel van dit project is het ontwikkelen van een analytisch hulpmiddel dat op basis van orderhistorie automatisch groepen producten identificeert die vaak samen worden gekocht. In plaats van te vertrouwen op categorie-structuren of intuïtieve aannames, zal het systeem gebruikmaken van echte verkoopdata om relaties tussen artikelen te detecteren.

Het project vormt de eerste fase van een bredere visie die gericht is op toekomstige optimalisatie van magazijn- en logistieke processen. In deze fase richten we ons nog niet op de fysieke plaatsing van producten in het magazijn of op simulaties van pickroutes. De focus ligt op het opbouwen van een solide databasis die inzicht geeft in welke producten daadwerkelijk samen voorkomen in orders en welke operationele impact zij hebben. Naast het creëren van clusters zal het systeem ook hun relatieve belang bepalen op basis van historische orderdata, wat de interpretatie van resultaten en het gebruik ervan in latere projectfasen vereenvoudigt.

Het systeem moet zo veel mogelijk autonoom functioneren — de gebruiker start de analyse en ontvangt een resultaat in de vorm van een toewijzing van producten aan clusters, eventueel met slechts enkele basisparameters. Gebruiksgemak betekent echter niet dat de technologie wordt vereenvoudigd. Het project gaat uit van het toepassen van geavanceerde methoden voor graafanaly-se en clustering om betrouwbare en bruikbare resultaten te verkrijgen.

Dergelijke oplossingen zijn geen experiment of nieuw idee. Analyse van producten die samen worden gekocht wordt al jarenlang toegepast in grote webshops en analytische systemen. De reden is eenvoudig — productcategorieën laten zelden zien hoe klanten daadwerkelijk winkelen. Dat iets in dezelfde categorie zit, betekent niet dat het samen wordt gekocht, en andersom geldt hetzelfde.

Waarom doen we dit? Bij duizenden SKU is het vrijwel onmogelijk om terugkerende patronen alleen op basis van observatie te herkennen. In de praktijk zijn veel magazijnbeslissingen nog steeds gebaseerd op intuïtie en aannames, omdat er geen hulpmiddel is dat echte relaties tussen producten zichtbaar maakt. Het automatisch creëren van clusters en het bepalen van hun belang binnen echte orders lost niet alle problemen op, maar biedt wel een objectiever referentiepunt. Hierdoor kunnen toekomstige logistieke en analytische beslissingen worden gebaseerd op data in plaats van uitsluitend op ervaring.

1. Scope en doelstellingen van het project

De scope van het project omvat het ontwikkelen van een tool die het mogelijk maakt om relaties tussen producten te analyseren op basis van orderhistorie en automatisch groepen artikelen te creëren die vaak samen worden gekocht. Een aanvullend onderdeel van de analyse is het bepalen van het relatieve belang van de ontstane clusters op basis van echte orderdata, zodat hun operationele betekenis beter kan worden begrepen. Het doel is om duidelijke en reproduceerbare resultaten te verkrijgen die in de toekomst kunnen worden gebruikt in vervolgprojecten, bijvoorbeeld gericht op magazijnoptimalisatie. Het systeem moet eenvoudig in gebruik zijn, maar tegelijkertijd gebaseerd op betrouwbare data-analyse en voorbereid zijn op verdere ontwikkeling. Vanaf het begin gaat het project uit van een aanpak waarbij kwaliteit, betrouwbaarheid en stabiliteit van de resultaten prioriteit hebben — de keuze van methoden en oplossingen wordt niet bepaald door technische eenvoud, maar door hun werkelijke analytische waarde.

Binnen het project dienen de volgende elementen te worden meegenomen:

- gebruik van echte orderdata als belangrijkste bron van informatie over relaties tussen producten,
- ontwikkelen van een methode voor het automatisch ophalen van inputdata via integratie met een van de bestaande systemen,
- creëren van één consistente en reproduceerbare manier om artikelen te groeperen in plaats van te vertrouwen op intuïtie of handmatige beslissingen,
- bepalen van het relatieve belang van clusters op basis van orderhistorie,
- gebruikmaken van de best beschikbare methoden en algoritmen om betrouwbare en stabiele analyseresultaten te verkrijgen,
- maximale vereenvoudiging van het gebruik — het starten van de analyse moet slechts minimale handelingen van de gebruiker vereisen,
- het systeem zo voorbereiden dat het in de toekomst met verschillende datasets kan werken en als basis kan dienen voor volgende projecten,
- onafhankelijkheid van de resultaten ten opzichte van bestaande productcategorieën,
- mogelijkheid om resultaten tussen verschillende tijdsperioden te vergelijken,
- veilige werking van de tool zonder invloed op lopende operationele systemen,
- presentatie van resultaten op een duidelijke en gemakkelijk te interpreteren manier,
- het beperken van extra technologische afhankelijkheden tot een minimum.

2. Vereenvoudigd werkingsschema van het systeem

Data

- ophalen van ordergegevens uit de geselecteerde periode in de vorm van rijen: order_id, order_date, sku_id, qty (optioneel),
- voorbereiden van een lijst met producten die aan elke order zijn gekoppeld,
- initiële opschoning van de data: orders met slechts 1 SKU uitsluiten en optioneel zeer grote orders beperken om ruis te verminderen.

Visueel: een lijst met orders waarbij elke order wordt behandeld als een verzameling producten, bijv.

Order 1001 → [SKU_A, SKU_B, SKU_C].

Bepalen van gezamenlijke voorkomens

- identificeren van producten die in dezelfde orders zijn voorgekomen,
- het tellen van alle SKU-paren die samen voorkomen.

Visueel: een tabel die laat zien hoe vaak twee producten samen voorkwamen, bijv.

SKU_A + SKU_B → 120 keer,
SKU_A + SKU_C → 15 keer.

Berekenen van de relatie-sterkte tussen SKU

- berekenen van de sterkte van relaties tussen producten met behulp van de Lift-maatstaf,
- wegfilteren van zwakke of toevallige relaties,
- voorbereiden van een productrelatie-grafiek.

Visueel: een netwerk van verbindingen tussen producten, waarbij dikkere lijnen producten voorstellen die vaker samen worden gekocht.

Clustervorming

- groeperen van producten op basis van de relatiegrafiek met behulp van het Louvain-algoritme,
- toewijzen van elk product aan een specifieke cluster_id.

Visueel: een lijst met productgroepen, bijv.

Cluster 1 → [SKU_A, SKU_B, SKU_D]
Cluster 2 → [SKU_X, SKU_Y]

Bepalen van het belang van clusters

- analyseren van orders om te bepalen hoe vaak producten uit specifieke clusters voorkomen,
- toekennen van een relatieve belangrijkheid aan clusters op basis van hun aandeel in orders.

Visueel: een samenvatting die het belang van clusters weergeeft, bijv.

Cluster 1 → 68% van de orders
Cluster 2 → 21% van de orders
Cluster 3 → 11% van de orders

3. Invoerdata

Om productclusters te kunnen creëren, is het noodzakelijk om basisinformatie over orders te verkrijgen. De minimale dataset moet zo eenvoudig mogelijk blijven, zodat de analyse stabiel en eenvoudig te onderhouden is.

Een belangrijke aanname van het project is maximale vereenvoudiging van het proces aan de gebruikerskant. Handmatige invoer van data is geen optie — bij duizenden SKU en een groot aantal orders zou dit onpraktisch en foutgevoelig zijn. Daarom moeten de gegevens automatisch worden aangeleverd via integratie met een systeem dat deze al bevat.

Vereiste data

- order_id — identificatie van de order,
- order_date — orderdatum (voor het bepalen van het analyse-tijdsbereik),
- sku_id — productidentificatie,
- qty (optioneel) — hoeveelheid van het product in de order.
  In de huidige fase is deze niet noodzakelijk voor het creëren van clusters, maar het is wel aan te raden om deze te verzamelen om de volledigheid van de data te behouden en toekomstige analyses mogelijk te maken.

Voorbereidende dataverwerking

- vóór de start van de analyse moeten orders met slechts één product (1-SKU) worden gefilterd, omdat ze geen informatie opleveren over relaties tussen producten en statistische berekeningen kunnen verstoren,
- elke order wordt behandeld als een verzameling unieke SKU (zonder duplicaten binnen dezelfde order).

Mogelijke databronnen

- **Optie A — integratie met WMS Locus (aanbevolen)**
  Het Locus-systeem beschikt over een archiefdatabase LTT waarmee een directe verbinding kan worden gemaakt om ordergegevens binnen een bepaalde periode op te halen. Deze aanpak maakt het mogelijk om precies de benodigde informatie te verkrijgen zonder extra tussenlagen en zonder uitbreiding van de bestaande infrastructuur. Door de eenvoud van implementatie en de directe beschikbaarheid van data is dit momenteel de beste oplossing.
- **Optie B — integratie met het ORP-systeem**
  Een alternatief is het koppelen van de applicatie aan het ORP-systeem. In dit geval moet de integratie worden ontwikkeld in samenwerking met Wolfpack. Dit kan leiden tot extra kosten, afhankelijkheid van externe partijen en mogelijk een langere doorlooptijd van het project.
- **Optie C — integratie met andere systemen (bijv. Yellowfin)**
  Het is theoretisch ook mogelijk om data te gebruiken uit andere analytische platforms, zoals Yellowfin. Dit vereist echter eerst een controle van de beschikbare integratiemethoden en een beoordeling of de data in een geschikte vorm beschikbaar is.

Aanbevolen aanpak

Omdat integratie met de LTT-database van het Locus-systeem directe toegang biedt tot de benodigde data en geen extra tussenlagen vereist, is dit op dit moment de meest rationele en snelste oplossing binnen het project.

4. Bepalen van gezamenlijke productvoorkomens

Na het ophalen van de orderdata is de volgende stap het bepalen welke producten daadwerkelijk samen voorkomen binnen dezelfde orders. In deze fase analyseert het systeem nog niet de sterkte van relaties en worden er nog geen clusters gevormd — het doel is uitsluitend om alle gezamenlijke voorkomens van producten te identificeren.

Elke order wordt behandeld als een lijst van SKU. Vervolgens wordt voor elke order een set productparen samengesteld die samen zijn voorgekomen. In de praktijk betekent dit dat wanneer een order drie producten bevat, alle mogelijke combinaties daarvan als potentiële relaties worden vastgelegd.

Voorbeeld:

Order 1001 → [SKU_A, SKU_B, SKU_C]

Op basis van deze order ontstaan de volgende paren:

SKU_A + SKU_B

SKU_A + SKU_C

SKU_B + SKU_C

Elke volgende order verhoogt de teller van deze paren. Als resultaat ontstaat een eenvoudige tabel die laat zien hoe vaak bepaalde producten samen zijn voorgekomen.

Visueel:

Productpaar	Aantal gezamenlijke voorkomens
SKU_A + SKU_B	120
SKU_A + SKU_C	15
SKU_B + SKU_D	87

In deze fase beoordelen we nog niet of een relatie sterk of toevallig is — we registreren uitsluitend feiten die voortkomen uit de orderdata. De resulterende lijst van gezamenlijke voorkomens vormt de basis voor de volgende stap, waarin de sterkte van relaties tussen producten wordt berekend.

5. Berekenen van de sterkte van productrelaties

Nadat is vastgesteld welke producten samen voorkomen in orders, is de volgende stap het bepalen hoe sterk deze relatie is. Het enkel tellen van gezamenlijke voorkomens is niet voldoende — populaire producten zullen met veel andere producten samen verschijnen, simpelweg omdat ze vaak worden gekocht. Daarom wordt de Lift-maatstaf gebruikt om de kwaliteit van relaties te beoordelen, waarna relatie-filtering wordt toegepast om alleen betrouwbare verbindingen te behouden.

Wat is Lift

Lift is een maatstaf die aangeeft of twee producten vaker samen worden gekocht dan op basis van toeval verwacht zou worden. Het helpt onderscheid te maken tussen situaties waarin producten daadwerkelijk een aankooprelatie hebben en situaties waarin ze alleen vaak in orders voorkomen omdat ze populair zijn. In de praktijk beantwoordt Lift de vraag:

Vergroot de aanwezigheid van product A in een order de kans dat product B ook voorkomt?

Formule:

$\text{Lift}(A,B) = P(A \cap B) / (P(A) * P(B))$

Praktische vorm:

$\text{Lift}(A,B) = (\text{together\_count} * \text{total\_orders}) / (\text{count\_A} * \text{count\_B})$

Relatiefiltering

Na het berekenen van de Lift-waarden wordt filtering toegepast met als doel toevallige of te zwakke relaties te verwijderen voordat de productgrafiek wordt opgebouwd.

In deze fase wordt in het bijzonder rekening gehouden met:

- een minimaal aantal gezamenlijke voorkomens van een productpaar (min_edge_support),
- een minimale Lift-drempelwaarde (lift_threshold).

De filtering heeft betrekking op relaties tussen producten en niet op de SKU zelf — hierdoor kunnen seizoens- of promotieclusters behouden blijven, terwijl de invloed van toevallige co-voorkomens door een klein aantal orders wordt beperkt.

Voorbeeldrelaties na het berekenen van Lift:

SKU-paar	Together Count	Lift
SKU_A + SKU_B	120	2.4
SKU_A + SKU_C	3	5.8
SKU_D + SKU_E	18	1.3
SKU_X + SKU_Y	1	9.5

Stel de volgende drempelwaarden vast:

• min_edge_support = 10

• lift_threshold = 1.2

Na filtering blijven alleen de relaties over die aan beide voorwaarden voldoen:

SKU-paar	Together Count	Lift
SKU_A + SKU_B	120	2.4
SKU_D + SKU_E	18	1.3

Relaties zoals SKU_A + SKU_C of SKU_X + SKU_Y worden afgewezen — ondanks een hoge Lift — omdat het aantal gezamenlijke voorkomens te laag is om ze als statistisch stabiel te beschouwen.

Interpretatie van de Lift-waarde

Lift ≈ 1 — geen betekenisvolle relatie,
Lift > 1 — producten worden vaker samen gekocht dan op basis van toeval,
Lift < 1 — producten komen minder vaak samen voor dan verwacht zou worden.

Voorbeeld:

Aantal orders: 10.000
SKU_A komt voor in 2000 orders
SKU_B komt voor in 1000 orders
SKU_A en SKU_B kwamen samen voor in 600 orders

Lift = (600 * 10.000) / (2000 * 1000)

Lift = 3

Een waarde van 3 betekent dat de producten drie keer vaker samen voorkomen dan op basis van toeval verwacht zou worden.

De manier waarop relaties op basis van Lift worden berekend en gefilterd heeft een directe invloed op de kwaliteit van de clusters in de volgende fase van de analyse. Daarom is het noodzakelijk om verschillende parameterwaarden te testen — zoals de Lift-drempel of het minimale aantal gezamenlijke voorkomens — en de meest optimale configuratie te kiezen. In sommige gevallen kan Lift alleen onvoldoende blijken; de analyse kan dan worden aangevuld met extra maatstaven, zoals Confidence of de Jaccard-index, om de stabiliteit en betrouwbaarheid van de resultaten te vergroten.

Na toepassing van de filtering ontstaat een set relaties met voldoende sterkte en statistische stabiliteit, die de basis vormt voor het opbouwen van de productgrafiek die in de volgende fase wordt gebruikt — het creëren van clusters.

.

6. Clustervorming van producten (Louvain)

Na het berekenen van de sterkte van relaties tussen producten is de volgende stap het groeperen ervan in grotere verzamelingen, oftewel clusters. Het doel is niet om nieuwe categorieën of kunstmatige indelingen te creëren, maar om natuurlijke productgroepen te ontdekken die rechtstreeks voortkomen uit orderdata.

Voor deze taak wordt het **Louvain-algoritme** gebruikt. Dit algoritme is ontwikkeld voor het analyseren van grote netwerken van relaties en wordt veel gebruikt binnen data science om natuurlijke clusters in data te detecteren.

In dit geval wordt het relatienetwerk gevormd door producten, en de verbindingen tussen hen zijn gebaseerd op de berekeningen uit de vorige fase (**Lift**). Elk product is een knooppunt in de grafiek, en de sterkte van de verbindingen weerspiegelt hoe vaak producten samen voorkomen in orders.

Hoe werkt het algoritme

Het algoritme analyseert de volledige structuur van relaties tegelijk en zoekt naar groepen producten waarin de interne verbindingen sterker zijn dan de verbindingen met de rest van het netwerk. Het proces verloopt iteratief:

- in het begin vormt elk product een aparte groep,
- het algoritme controleert of het verplaatsen van een product naar een andere groep de kwaliteit van de indeling verbetert,
- producten met sterke onderlinge relaties beginnen clusters te vormen,
- de structuur wordt vereenvoudigd en opnieuw geanalyseerd,
- het proces stopt wanneer verdere wijzigingen het resultaat niet meer verbeteren.

Het aantal clusters wordt niet handmatig bepaald — het komt rechtstreeks voort uit de data.

Toepassing in logistiek en magazijnsystemen

Een aanpak gebaseerd op analyse van relaties tussen SKU is geen nieuw concept en wordt toegepast in moderne logistieke systemen, vooral op het gebied van slotting-optimalisatie en analyse van ordergedrag.

Voorbeelden van systemen die vergelijkbare concepten gebruiken:

- Manhattan Associates WMS — optimalisatie van productplaatsing op basis van orderrelaties,
- Blue Yonder (JDA) — analyse van SKU-relaties en aanbevelingen voor magazijnlayout,
- SAP EWM — geavanceerde slottingstrategieën die rekening houden met de frequentie van gezamenlijke productvoorkomens,
- 3PL- en e-commerceoplossingen die gebruikmaken van zogenaamde order affinity of basket analysis.

In veel magazijnen zijn dit soort analyses verborgen binnen systemen als onderdeel van grotere optimalisatiemodules. Het voorgestelde hulpmiddel realiseert hetzelfde concept op een meer directe manier — door een onafhankelijke analytische laag te creëren die uitsluitend is gebaseerd op orderdata.

Voorbeeldresultaat

Het resultaat van het algoritme is dat elk product wordt toegewezen aan een specifiek cluster:

Sku Id	Cluster Id
SKU_001	1
SKU_002	1
SKU_003	1
SKU_014	2
SKU_015	2
SKU_021	3

Dit kan ook worden weergegeven als een overzicht van groepen:

Klaster	Sku
Klaster 1	SKU_001, SKU_002, SKU_003
Klaster 2	SKU_014, SKU_015
Klaster 3	SKU_021

Het resultaat van deze fase is een gestructureerde lijst van sku_id → cluster_id. De verkregen clusters vormen een georganiseerde analytische laag die de werkelijke relaties tussen producten beschrijft en dienen als uitgangspunt voor volgende analysetappen, zoals het bepalen van het belang van afzonderlijke clusters of het verder benutten van de data in toekomstige projecten.

7. Bepalen van het belang van clusters

Na het vormen van clusters kan hun werkelijke operationele belang worden bepaald op basis van de orderregelgeschiedenis. Het doel van deze fase is niet om de clusterstructuur te wijzigen, maar om vast te stellen welke productgroepen de grootste belasting veroorzaken binnen het pickproces.

Hoe de analyse werkt

- elke orderregel wordt aan een cluster toegewezen op basis van sku_id,
- het totale aantal orderregels dat tot elk cluster behoort wordt geteld,
- het aandeel van de orderregels van elk cluster ten opzichte van alle orderregels in de geanalyseerde periode wordt berekend,
- dit aandeel wordt beschouwd als een maat voor de relatieve belangrijkheid van het cluster — hoe groter het aandeel, hoe vaker het cluster betrokken is bij pickactiviteiten en hoe groter de operationele belasting die het veroorzaakt.

Voorbeeld:

Orderregels:

Order 1001 → SKU_A (Cluster 1), SKU_B (Cluster 1), SKU_X (Cluster 3)
Order 1002 → SKU_C (Cluster 1), SKU_Y (Cluster 2)

Telling van orderregels:

Cluster 1 → 3 regels
Cluster 2 → 1 regel
Cluster 3 → 1 regel
Totaal → 5 regels

Samenvatting van het belang van clusters:

Cluster 1 → 60% aandeel in orderregels
Cluster 2 → 20%
Cluster 3 → 20%

Interpretatie

- een hoger percentage geeft een groter werkelijk gebruik van het cluster binnen magazijnoperaties aan,
- de metriek weerspiegelt de frequentie van pickactiviteiten in plaats van alleen de aanwezigheid van een cluster in orders,
- analyse op orderregelniveau vermindert de invloed van individuele zeer populaire SKU op het eindresultaat.

In deze fase betreft het een aanvullende analyse die geen invloed heeft op het clusteringproces zelf, maar helpt om de verkregen resultaten beter te interpreteren.

8. Mogelijke toekomstige toepassingen in het magazijn

Het huidige project eindigt bij het creëren van productclusters en het bepalen van hun relatieve belangrijkheid. In deze fase grijpt het systeem niet in op de magazijnlayout of op lopende operationele processen. De verkregen data vormt echter een solide basis voor verdere analyses en mogelijke verbeteringen in de toekomst.

Belangrijkste mogelijke richtingen voor het gebruik van de resultaten:

- **Ondersteuning bij optimalisatie van magazijnlayout en pickroutes**
  Clusters laten zien welke producten daadwerkelijk samen voorkomen in orders, wat beslissingen kan ondersteunen rond SKU-plaatsing, het ontwerp van operationele zones en het verkorten van loopafstanden — zowel bij single-order picking als bij batch picking.
- **Optimalisatie van replenishment**
  Clusteranalyse kan helpen bij het identificeren van producten die vaak samen worden verbruikt. Hierdoor wordt het mogelijk om:
    - o replenishment te plannen voor volledige productgroepen in plaats van individuele SKU,
    - o situaties te beperken waarin pickers meerdere lege locaties tegenkomen binnen dezelfde route,
    - o de volgorde van replenishment beter te prioriteren op basis van clusterbelang.
- **Objectivering van operationele beslissingen**
  In plaats van uitsluitend op intuïtie te vertrouwen, kunnen geplande wijzigingen worden vergeleken met echte data over productrelaties.
- **Basis voor toekomstige simulaties en optimalisatietools**
  Clusters en hun belangrijkheid kunnen worden gebruikt in volgende projecten, bijvoorbeeld voor het simuleren van verschillende layoutscenario's of het analyseren van de impact van veranderingen op het pickproces.
- **Detectie van mogelijk suboptimale layoutbeslissingen**
  Clusteranalyse maakt het mogelijk situaties te identificeren waarin producten die vaak samen worden besteld fysiek ver uit elkaar liggen in het magazijn. Dit kan een signaal zijn om SKU-plaatsing opnieuw te evalueren en helpt bij beter onderbouwde layoutbeslissingen op basis van operationele data.

Het is belangrijk te benadrukken dat dit potentiële toepassingen zijn die voortkomen uit de analytische mogelijkheden van het project. Het doorvoeren van concrete wijzigingen in het magazijn zou een aparte projectfase en aanvullende operationele analyses vereisen.