> SPSS Magazine ^s

EEN UITGAVE VAN **SPSS** BELUX JAARGANG 1 NR 1 - MAART 2005 INFORMATIE VOOR PROFESSIONALS IN ONDERZOEK, ANALYSE & REPORTING



Clementine 9.0: nog sneller meetbare ROI op uw data mining proces

Clementine staat al bekend in de markt als dé data mining workbench die een meetbare ROI oplevert binnen afzienbare tijd. In Clementine 9.0 wordt dit voortgezet en door feedback van klanten is de productiviteit voor het maken van modellen verbeterd en kunt u het maximale uit uw database/datawarehouse halen.

In-database 'minen'

Met deze nieuwste versie van Clementine haalt u nog meer uit uw database, want SPSS integreert de data mining faciliteiten van grote namen uit de database wereld met de gemakkelijke interface van Clementine. Dit in-database 'minen' is mogelijk met SQL Server, IBM DB2 Intelligent Miner en Oracle 10 Data Mining. Met Clementine kunt u modellen bouwen en scoren in deze database. Met deze laatste
mogelijkheid is Clementine uniek om het gehele data mining proces, van data preparatie tot scoring en deployment, in de database omgeving uit te voeren en biedt Clementine een zeer korte 'time-tosolution'.

Daarnaast maakt Clementine het mogelijk om tot een uitgebreide oplossing te komen: data mining wordt ingezet als een algeheel proces binnen de organisatie. De add-on voor een beheersysteem en integratie van text mining ondersteunen de organisatiebrede inzet van het data mining proces.

Colofon

Coördinatie Dafne Brouwer

Realisatie EVV Reclamestudio

Aan dit nummer werkten mee Rob Geuverink (Ahrend Office Products) Jos den Ronden (SPSS) Wiepke Van der Wal (SPSS) Sophie Van Malder (SPSS)

SPSS magazine wordt kostenloos verspreid onder relaties van SPSS Belux. Artikelen mogen worden overgenomen na schriftelijke toestemming van de redactie. Voor meer informatie over SPSS software kunt u zich richten tot de afdeling sales. En in geval van technische vragen kunt u zich wenden tot de helpdesk.

SPSS is een geregistreerd handelsmerk en de overige genoemde SPSS software zijn handelsmerken van SPSS Inc. Overige producten en applicaties zijn handelsmerken van respectievelijke organisaties.

SPSS Belux

Buro & Design Center B38 Heizel Esplanade Heysel 1020 Brussels Tel. : +32 2 474 0360 Fax : +32 2 474 0363 Email : **belgium@spss.com** Internet : **www.spss.com/be** Ook kan Clementine ingezet worden voor delen van het gehele proces.

Beheersysteem

Met het nieuwe SPSS Model Management systeem is het mogelijk om objecten gemaakt in Clementine -zoals streams, modellen en output- op te slaan in een beheersysteem en te delen met anderen. Een groot voordeel is dat objecten voor meerdere mensen in de organisatie toegankelijk zijn, waardoor het wiel niet opnieuw uitgevonden hoeft te worden. Daarnaast is het beveiligd en er kan aangegeven worden welke groepen toegang hebben tot welke informatie. Uiteraard bevat dit beheersysteem ook een uitgebreide terugzoekfunctie.

Text mining als add-on direct inzetbaar

Text mining roept u in deze versie direct aan vanuit de interface. Dus u kunt ook ongestructureerde tekstdata direct structureren en zo uw dataset verrijken met concepten uit bijv. gespreksverslagen, emails, bezoekverslagen en elke andere relevante bron van ongestructuureerde teksten. Een telecom bedrijf heeft -door het integreren van gespreksverslagen uit het call-center in het analyse proces- zijn voorspellingen welke klanten de grootste kans hebben om hun abonnement op te zeggen met 20% verbeterd. Door deze klanten op de juiste wijze te benaderen is dit een jaarlijkse besparing van enkele miljoenen euro's.

Interactieve beslisbomen

Met Clementine 9.0 kunt u verschillende modellen testen en snel tot de conclusie komen welke het beste is. Deze functionaliteit voor beslisbomen is uitgebreid met 3 nieuwe algoritmen en een verbeterde interactiviteit. De nieuwe algoritmen zijn: Quest, CHAID en exhaustive CHAID. Dit geeft de breedste range aan mogelijkheden. Verder vindt u een scala aan mogelijkheden om uw beslisboom op uw manier samen te stellen. Naast standaard opties is het mogelijk om zelf aan te geven hoe u de beslisboom of een tak wilt laten groeien. Wilt u verschillende mogelijkheden evalueren tijdens dit proces? Met het maken van grafieken heeft u direct een duidelijk beeld van de resultaten. Zijn de resultaten dusdanig dat u een model wilt maken waarmee u verder wilt werken? U schikt dit model direct in uw data mining proces.

Clementine inzetten in uw organisatie

Wilt u weten wat Clementine voor u kan betekenen? Fax dan het **antwoordformulier** naar +32 2 474 0363 voor meer informatie. Wilt u er zelf mee werken, dan kunt u **de eendaagse training 'Kennismaking met Clementine'** bijwonen (data vindt u in het trainingsschema).



Figuur: De evaluatie grafiek laat u zien hoe goed het data mining model (\$C-pep) in dit geval respons kan voorspellen. De onderste as laat de totale populatie zien, de verticale as (gain) laat het totaal aantal respondenten zien. De zwarte lijn is de zgn. '50/50' lijn: als u zou moeten gokken wie van uw populatie gaat responderen zou u bij het benaderen 50% van de populatie ook ongeveer 50% van de respondenten vinden. Het data mining model kan, in de testset, 88% van alle respondenten vinden in 51% van de populatie. U hoeft dus niet 37% van uw populatie te benaderen die u anders wel had moeten benaderen om aan de 88% respondenten te komen. Afhankelijk van de aantallen is dit een flinke besparing. Daarnaast ziet u in de grafiek de blauwe lijn, met het label 'rule': dit is de huidige business rule (het zgn. onderbuik gevoel) die deze organisatie tot nu toe had gebruikt voor de selectie van haar doelgroep. Nu afgezet tegen het data mining model ziet u dat dit onderbuik gevoel het nauwelijks beter doet dan gokken...

Text mining: een kritisch traject voor beter zakelijk inzicht

De zoektocht naar het volledig inzicht in de bedrijfsactiviteiten en klanten van een onderneming heeft tot aanzienlijke veranderingen op het gebied van data warehousing-initiatieven geleid.

In hun streven naar dit volledig inzicht in bedrijfsactiviteiten hebben organisaties nauwgezet hun transactionele, operationele en concurrentie-informatie naar 'gestructureerde vormen' geconverteerd. Om deze gegevens vervolgens vast te leggen zodat met business intelligence tools, zoals predictive analytics en OLAP, de analyse kan beginnen en de resultaten kunnen worden gerapporteerd. Het vastleggen en analyseren van gestructureerde gegevens is, echter, slechts één aspect van de zoektocht naar beter zakelijk inzicht. Een aspect van business intelligence dat vaak ononderzocht blijft is tekstuele of 'ongestructuurde' informatie.

Analyse van tekst blijkt waardevol

Waardevolle bedrijfsinformatie zit opgesloten in emails, statusmemo's, bezoekrapporten, gereguleerde verslagen en in een verscheidenheid aan andere documenten. De informatie die tekst bevat, is een onaangesproken bron van business intelligence. Dergelijke informatie geeft vaak de achtergrondkennis en het inzicht die nodig zijn voor het nemen van zakelijke beslissingen in bepaalde situaties. Gekoppeld aan de informatie die aan gestructureerde gegevens worden onttrokken, kan tekstanalyse organisaties de context leveren die nodig is om de juiste beslissingen te nemen.

De reden waarom tekst niet in het business intelligence-systeem is opgenomen is dat, tot voor kort, het moeilijk was om de analyse van tekst te automatiseren. Er bestaan echter tegenwoordig text mining technologieën die het mogelijk maken om tekstuele gegevens op efficiëntere wijze te onderzoeken om zodoende beter inzicht te krijgen. Met behulp van text mining technieken worden verzamelingen van tekstueel materiaal geanalyseerd waardoor essentiële concepten, verborgen thema's, relaties en trends kunnen worden blootgelegd.

Een inleiding in text mining

Naast de manuele benadering - waarbij mensen documenten uitkammen in een poging om de relevante informatie uit tekst op te diepen - bestaan er andere manieren van text mining, die veel efficiënter en effectiever zijn. Een benadering is die van de geautomatiseerde oplossing, die gebaseerd is op statistiek en neurale netwerken.

Door de inzet van computertechnologie kunnen wezenlijke concepten sneller worden gescand en gecategoriseerd. Helaas is de nauwkeurigheid van dergelijke oplossingen tamelijk laag. De meeste op statistiek gebaseerde systemen tellen simpelweg het aantal keren dat termen of begrippen voorkomen en berekenen hun statistische verwantschap met gerelateerde termen of begrippen. Zodoende worden er veel irrelevante resultaten geproduceerd (ook wel bekend als ruis) en blijven bepaalde uitkomsten die wel gevonden hadden moeten worden (bekend onder de term 'stilte') onontdekt.

Een ander type text mining is de op linguïstiek gebaseerde text mining. Deze vorm van text mining biedt de snelheid en rendabiliteit van statistisch georiënteerde systemen. Ook biedt het een veel grotere nauwkeurigheidsgraad, waarbij een veel geringere menselijke interventie vereist is. De op linguïstiek gebaseerde text mining technieken komen voort uit het onderzoeksveld dat bekend staat als Natural Language Processing (NLP), ook wel bekend onder de naam computerlinguïstiek.

Om het verschil tussen statistisch-georiënteerde en linguïstisch-georiënteerde benaderingen duidelijk te maken een voorbeeld. Een statistisch-georiënteerde oplossing bevat waarschijnlijk een regel die zegt dat indien een tekst over 'zakelijk leiderschap' gaat, het woord 'manager' waarschijnlijk ook in deze tekst zal worden gebruikt. Indien beide termen voorkomen zal het statistisch-georiënteerde systeem het document selecteren. Dit betekent, echter, dat de gescande teksten lang genoeg moeten zijn om als bron te fungeren en zelfs dan is er geen garantie dat de verwachte wijze waarop het concept wordt uitgedrukt in de tekst zal voorkomen. Een onderzoeker kan zodoende een interessant artikel over leiderschap over het hoofd zien, indien de auteur de term 'leidinggevend persoon' in plaats van 'manager' heeft gebruikt. Een taalkundige benadering vergelijkt termen als 'leidinggevend persoon', 'manager', en zelfs 'mngr' met elkaar indien ze in soortgelijke contexten worden gebruikt. Bovendien kunnen linguïstisch-georiënteerde oplossingen de toon van de tekst analyseren, en het taalbegrip maakt korte metten met de ambiguïteit die tekst eigen is.

De combinatie van text mining met data mining

Indien data mining met text mining wordt gecombineerd, wint text mining aan voorspellend vermogen. Alle voorspellende technieken die beschikbaar zijn door data mining kan men toepassen op concepten die uit tekstuele gegevens verkregen zijn. Deze concepten worden geïntegreerd met de bestaande gestructureerde gegevens en het totale gegevensbestand wordt geëvalueerd met behulp van data mining.

De combinatie van text mining en data mining kan worden toegepast om een groot scala aan business issues in vele bedrijfstakken aan te pakken. Een telecommunicatiebedrijf, bijvoorbeeld, kan deze combinatietechniek gebruiken om het probleem van klantenontrouw aan te pakken. In dit type organisatie worden gegevens over organisatie, klant, demografie, gebruik en andere gestructureerde gegevens door een data mining tool gebruikt om de geneigdheid van de klant om te vertrekken te voorspellen. Echter, aangezien de redenen voor overlopen van de klant vaak aan call centers of in enquêtes worden verwoord, zou text mining een oplossing zijn om de verbale respons van de klant te analyseren. Hierdoor identificeert men de belangrijkste concepten - slechte service, te duur, onvoldoende bereik - waarna de concepten als toegevoegde velden worden gestructureerd, die aan de analyse kunnen worden toegevoegd. Data mining wordt dan gebruikt om de waarschijnlijkheid van de volgende actie van de klant te clusteren, classificeren, segmenteren, profileren en voorspellen.

Een rijke informatiebron aanboren

Tekst is een onaangeboorde, rijke informatiebron, en een essentieel element van de business intelligence mix. Door text mining en data mining in hun business intelligence-systemen te incorporeren, zullen organisaties in de gelegenheid worden gesteld om meer begrip van hun bedrijfstak en van hun klanten te krijgen.

Meer informatie?

Wilt u meer weten over het begrip text mining en hoe u dit in kunt zetten binnen uw organisatie? Fax dan het **antwoordformulier** retour. Of neem contact op met onze afdeling sales op telefoonnummer +32 2 474 0360.

Tips & Tricks: combineren van bestanden (1)

Dit keer besteden we aandacht aan het combineren van bestanden binnen SPSS. Hierin zijn verschillende varianten denkbaar. Hieronder zullen deze varianten worden uitgewerkt. Naast de dialoogvensters laten we ook de syntax zien, omdat er in de syntax meer functionaliteit aanwezig is.

Aangezien er veel verteld kan worden over het combineren van bestanden is dit de eerste van de twee *Tips & Tricks* die aan dit onderwerp is gewijd.

Toevoegen cases (gelijke variabelen)

Situatie (links: deze bestanden hebben we; rechts: het gewenste bestand).



Toelichting:

4

Beide bestanden bevatten exact dezelfde variabelen.

Procedure:

- Open één van de beide bestanden, bijvoorbeeld *leerlingen* klas 1 wiskunde.sav.
- Kies: Data | Merge Files | Add Cases.
- Selecteer het tweede bestand, hier: *leerlingen klas 2 wiskunde.sav.*

U krijgt het volgende dialoogvenster:



Toelichting:

- Onder *Unpaired Variables* staan variabelen die wel aanwezig zijn in het ene bestand, maar niet in het andere bestand. Omdat in ons voorbeeld de variabelen in beide bestanden identiek zijn is dit leeg.
- Onder Variables in New Working Data File staan de variabelen die in het gecombineerde bestand zullen verschijnen.
- Met de optie Indicate case source as variable kunt u besluiten om een variabele aan te laten maken die aangeeft uit welk bestand de case komt. We vinken deze optie aan.

Als we met OK bevestigen dan krijgen we het volgende bestand:

eerling		Anne			
	leerling	klas	wiskunde	source01	Va
1	Anne	1	9	0	
2	Fred	1	4	0	
3	Lotte	1	7	0	
4	Vincent	1	8	0	
5	Corine	2	9	1	
6	Edwin	2	6	1	
7	Fred	2	3	1	
8	Joep	2	7	1	
9					-

Merk op, dat cases uit het werkbestand (het geopende bestand, hier *leerlingen klas 1 wiskunde.sav*) de waarde 0 krijgen op *sourceo1*, cases uit het bestand dat bij *Add Cases* is aangegeven krijgen de waarde 1. In syntax krijgt dit de volgende vorm:

SALENCE - 21-22 SALENCE
Edit Yeew Analyze Graphic Utilities Bun Window Help
DD FILES /FILE=* /FILE=*C:\cursus\leerlingen klas 2 wiskunde.sav' /IN=source01.
ARIABLE LABELS source01 'Case source is C:\cursus\leerlingen klas 2 wiskunde.sav'

KECUTE.

Toelichting:

- Bij de deelopdracht /FILE wordt elk van de bestanden aangegeven. Het sterretje * verwijst naar het werkbestand (het geopende bestand leerlingen klas 1 wiskunde.sav).
- Bij de deelopdracht //N wordt een variabelennaam opgegeven die de bron aangeeft.
- Met de opdracht VARIABLE LABELS krijgt sourceo1 een duidelijke omschrijving.

Belangrijk voordeel van syntax is, dat er meerdere bestanden in één keer kunnen worden gecombineerd, dit door naar elk van de bestanden te verwijzen met de deelopdracht /*FILE*. Stel bijvoorbeeld, dat u 5 bestanden wilt combineren dan zou dat via de dialoogvensters in 4 keer moeten, terwijl dat met syntax in één keer kan.

Toevoegen van variabelen (gelijke cases!)

Situatie (boven: de twee bestanden die beschikbaar zijn; onder: het gewenste bestand).



1 - Incoloro	200 000 20	(Acres	a a a a	The state	2.4
· nemes	learing	idas	wiskunde	frans	var -
1	Anne	1	9	9	_
2	Fred	1	4	3	_
3	Lotte	1	7	10	_
4	Vincent	1	8	9	

We merken duidelijkshalve op, dat beide bestanden exact <u>dezelfde cases</u> bevatten, in exact <u>dezelfde volgorde</u>. We kunnen de bestanden dus combineren door ze botweg tegen elkaar te leggen.

Procedure:

- Open één van de beide bestanden, bijvoorbeeld *leerlingen* klas 1 wiskunde.sav.
- Kies: Data | Merge Files | Add Variables.
- Selecteer het tweede bestand, hier: *leerlingen klas 1 frans.sav.*

U krijgt het volgende dialoogvenster:



Toelichting:

- Onder *Excluded Variables* staan variabelen die niet mee zullen worden genomen naar het nieuwe bestand. In dit geval zijn dat de variabelen *klas* en *leerling*, afkomstig uit het *leerlingen klas* 1 *frans.sav*, getuige de + (onderaan in het dialoogvenster ziet u dat de + refereert aan dat bestand; het * refereert aan het werkbestand, ofwel het geopende bestand).
- Onder New Working Data File staan de variabelen genoemd die in het gecombineerde bestand zullen verschijnen. Tevens ziet u achter elke variabele uit welk bestand de variabele komt.

 Ook hier kunt u besluiten om met de optie Indicate case source as variable een variabele te laten aanmaken die aangeeft uit welk bestand de case komt. Deze optie zetten we aan.

Als we met OK bevestigen dan krijgen we het volgende bestand:

: leerling		A	nne			
	leerling	klas	wiskunde	frans	source01	
1	Anne	1	9	9	1	
2	Fred	1	4	3	1	
3	Lotte	1	7	10	1	
4	Vincent	1	8	9	1	
5						

Merk op, dat elke case de waarde 1 heeft op *sourceo1*. Hiermee wordt aangegeven dat elke case ook in het externe bestand *leerlingen klas 1 frans.sav* aanwezig is.

In syntax krijgt dit de volgende vorm:

Syntax2 - SPSS Syntax Editor	
Sle Edit Yew Analyze Graphs Utilities Bun Window Help	
BRA 5 0 10 10 10 10 10	
	-
MATCH FILES	
/FILE=*	
/FILE='C:\cursus\leerlingen klas 1	frans.sav'
/RENAME (klas leerling = d0 d1)	
/IN=source01	
/DROP= d0 d1.	
VARIABLE LABELS source01	
<pre>'Case source is C:\cursus\leerlinge EXECUTE.</pre>	en klas 1 frans.sav*
	·
SPS5 Processor	s ready

Toelichting:

- De deelopdrachten /FILE en /IN hebben dezelfde functie als hiervoor beschreven bij ADD FILES.
- Bij /RENAME worden de variabelen klas en leerling uit leerlingen klas 1 frans.sav herbenoemd tot respectievelijk do en d1.
- Middels /DROP worden do en d1 (feitelijk dus klas en leerling uit leerlingen klas 1 frans.sav) verwijderd.

Voordeel van syntax ten opzichte van menu is ook hier weer, dat er meerdere databestanden in 1 keer kunnen worden gecombineerd.

Nogmaals dient te worden opgemerkt, dat de cases in beide bestanden dezelfde zijn, en bovendien in dezelfde volgorde staan. Deze situatie zal wellicht niet vaak voorkomen. Het zal vaker voorkomen, dat de cases in beide bestanden niet dezelfde zijn, zie het vervolg.

Toevoegen van variabelen (ongelijke cases en/of variabelen)

Situatie (boven: de twee bestanden die beschikbaar zijn; onder: het gewenste bestand).

leering			Anne			1 : leering			Anne	
	learling	klas	wiskunde	-			leerling	kla	tekenen	-
1	Anne	1	9			1	Anne	1	9	-
2	Fred	1	4			2	Fred	1	6	
3	Lotte	1	7		+	3	Maaike	1	9	
4	Vincen	1	8			4	Dyane	2	9	
5	Corine	2	9			5	Fred	2	3	
6	Fred	2	3			6	Onno	2	3	
7	Joep	2	7			7	-	_		
> \ Dat	a View K	on a		C		• • Dat	a View X V	1		

eering			Aree			
	learling	kla	wiskunde	tekenen		
- 1	Anne	1	9	9		
2	Fred	1	4	6		
3	Lotte	1	7			
4	Maaike	1		9		
5	Vincen	1	8			
6	Corine	2	9			
7	Dyane	2		9		
8	Fred	2	3	3		
9	Joep	2	7			
10	Onno	2		3		

We merken op:

- In de bestanden zitten andere cases én variabelen.
- De bestanden zijn gesorteerd op klas en daarbinnen op leerling.
- Er zit zowel in klas 1 als 2 een Fred.

We kunnen de bestanden nu niet botweg tegen elkaar leggen, zoals in de voorgaande situatie. We zullen beide bestanden moeten combineren op een sleutel: wiskunde van Anne moet worden gelegd tegen tekenen van Anne. Of ook: Fred uit klas 1 in het ene bestand moet worden samengenomen met Fred uit klas 1 in het ander bestand. We zien dus, dat de sleutel bestaat uit de combinatie van klas en leerling.

Procedure:

- Sorteer beide bestanden op klas, daarbinnen op leerling. Open één van de beide bestanden, bijvoorbeeld leerlingen
- wiskunde.sav.
- Kies: Data | Merge Files | Add Variables.
- Selecteer het tweede bestand, hier: leerlingen tekenen.sav.



U krijgt het volgende dialoogvenster:

Add Variables fromette\leerlingen tekenen	i.sav	×
Excluded Variables: klas (+) leerlingc (+)	New Working Data File: leerlingc (*) klas (*) wiskunde (*) tekenen (+)	OK Paste Reset Cancel Help
Hensme Match cases on key variables in sotted files Both files provide cases Egternal file is keyed table Working Data File is keyed table	Key Variables:	
Image: Indicate case source as variable: source01 (*) = Working Data File (+) =ette\leerlingen teke	enen.sav	

Toelichting:

- Onder Excluded Variables staan klas en leerling uit leerlingen tekenen.sav, getuige de +.
- Onder New Working Data File staan de variabelen genoemd die in het gecombineerde bestand zullen verschijnen. Tevens ziet u achter elke variabele uit welk bestand de variabele komt.
- Ook hier kunt u besluiten om met de optie Indicate case source as variable een variabele te laten aanmaken die aangeeft uit welk bestand de case komt. Deze optie zetten we aan.

Het is nu zaak om aan te geven dat klas en leerling de sleutel

- Vink de optie Match cases on key variables in sorted files aan. • Selecteer klas onder Excluded Variables: en klik die naar Key
- Selecteer leerling onder Excluded Variables: en klik die naar

Als we met OK bevestigen dan krijgen we het volgende bestand:

leering			Anne		
	leerling	klas	wiskunde	tekenen	source01
1	Anne	1	9	9	1
2	Fred	1	4	6	1
3	Lotte	1	7	14	0
4	Maaike	1		9	1
5	Vincen	1	8	14	0
6	Corine	2	9		0
7	Dyane	2		9	1
8	Fred	2	3	3	1
9	Joep	2	7		0
10	Onno	2		3	1

Toelichting:

- Het nieuwe bestand bestaat uit alle cases en alle variabelen.
- Daar waar gegevens onbekend zijn (bijvoorbeeld Lotte heeft
- geen cijfer voor tekenen), daar komen system missings.

Uiteraard kan ook hier gebruik gemaakt worden van syntax.

		-	_
MATCH F	ILES		
/FILE	= 'C:\cursus\leerlingen tekenen.sav'		5
/IN	= source01		
and the second			
VARIABL	E LABELS source01		
'Case	source is C:\Cursus\leerlingen tekenen.sav'.	100	
BABCUIS	5		

Conclusie

In het voorgaande zijn drie situaties besproken om bestanden te combineren:

- Toevoegen van cases, bij gelijke variabelen.
- Toevoegen van variabelen, bij gelijke cases.
- Toevoegen van variabelen én cases.

De volgende keer besteden we aandacht aan de fijne nuances die hierin kunnen worden aangebracht.

Training

Wilt u meer weten over syntax en combineren van bestanden, dan bevelen wij de training 'SPSS Advanced, Further Data Manipulation' warm aan. Hier wordt bijvoorbeeld ook aandacht besteed aan foutboodschappen die u bij het combineren van bestanden kunt krijgen. Voor meer informatie kunt u contact opnemen met de afdeling sales op telefoonnummer +32 2 474 0360. Voor inschrijving faxt u het antwoordformulier retour.

Ahrend ziet haar concurrentiepositie en marge verbeteren

Uitdaging

Ahrend is in 1896 in Amsterdam opgericht door Jacobus Ahrend. Ahrend is nu uitgegroeid tot een internationale topspeler op het gebied van kantoorinrichting en -artikelen. In 2002 had Ahrend een omzet van € 471,5 mln., 1.756 medewerkers en kantoren in Nederland, België, Duitsland, Tsjechië en Verenigd Koninkrijk.

Uitdaging

In de markt van kantoorartikelen staan vooral de marges onder druk. Om in deze markt slagvaardig op te kunnen treden, zijn flexibiliteit en kwaliteit van belang.

Voor 'business users' van afdelingen als Marketing &

e-Commerce, Verkoop, Financiën en Inkoop was het moeilijk om de informatie die zij nodig hadden snel te verkrijgen. Alle informatie was te vinden in het ERPsysteem BRIDGE/400 van kaertSoftware. Echter voor flexibele en snel toegankelijke management informatie was een data warehouse noodzakelijk.

Oplossing

Sinds 1998 werkt Ahrend Office Products International met de business intelligence (BI) oplossing ShowCase Suite from SPSS (SPSS). De SPSS oplossing bestaat uit: Warehouse Builder als ETL-tool voor het inrichten van een data warehouse en het laden van de OLAP kubussen, te weten Essbase, Analyzer als front-end oplossing voor online analyses, Query/Report Writer



voor standaard rapportages en Warehouse Manager voor het beheer van de omgeving (zoals bijv. autorisatie en resource management). Momenteel maken 100-120 werknemers gebruik van SPSS. Volgens Rob Geuverink, Applicatiespecialist bij Ahrend Office Products, levert SPSS al geruime tijd een significante bijdrage aan het bedrijfsresultaat van Ahrend.

Vandaag detailinzicht in de omzet en marges van gisteren

Marco Drijver, Sales Analist Nederland, gebruikt SPSS om gegevens snel en gemakkelijk uit het systeem te halen. Hij concentreert zich vooral op marges, omzetten en prijsverschillen. Zijn taak is om te kijken of er afwijkingen zitten in de cijfers met bijbehorende oorzaken. "Iedere ochtend bekijk ik in een dagkubus, een overzicht met de omzet van de dag, of er afwijkingen te vinden zijn. Hier wordt dan direct actie op ondernomen en de gegevens worden meteen hersteld. Vroeger kwam men er pas na een maand of langer achter en dan kostte het veel meer tijd om de oorzaak nog te achterhalen," aldus Drijver, "ook worden alle data al in de juiste vorm opgehaald, zodat er niets samengevoegd hoeft te worden. Waar ik voorheen 3 dagen mee bezig was, doe ik nu in enkele uren."

Monitoren van budgetten in één oogopslag

In de financiële kubus kan per afdeling vergeleken worden hoe de kosten zich verhouden tot het budget. Voorheen werden de kosten geprint via 'print screen' en ging men om de tafel zitten met de verschillende afdelingen. Met SPSS is het voordeel dat Financiën de budgetten ten allen tijde kan monitoren en in kan springen als een afdeling of kostenpost uit de pas loopt.

Met 1 druk op de knop weten hoe de klanten 6 ervoor staan

Account managers willen liever klanten bezoeken dan hun tijd besteden aan analyses. "Wat het grote voordeel is van SPSS, is dat ik met één druk op de knop direct zie welke klant achterloopt qua omzet vergeleken met vorig jaar of van welke productgroep hij meer zou moeten bestellen als je kijkt naar de algemene tendens. Dit scheelt mij ontzettend veel tijd, want nu kan ik alles in de 'standaardviews' zien," aldus Susanne Klabbers, Account Manager. Analyses worden losgelaten op individuele klanten. Als er negatieve uitschieters zijn laat Analyzer, de frontend-tool, een rood stoplicht zien, zodat de account manager weet waar de aandacht op gevestigd moet worden. In de toekomst wil men kijken hoe die verhoudingen per marktsegment liggen en dit toepassen op bedrijven uit dit marktsegment. Kortom, met SPSS wordt tijdwinst geboekt voor zowel de voorbereiding op een komend gesprek als het pro-actief reageren op actuele zaken.

Het inkoopbeleid verloopt effectiever en efficiënter

De inkoopafdeling van Ahrend is georganiseerd als matrix organisatie. Dit betekent dat de verschillende productgroepen verdeeld zijn over de inkoopmanagers van Nederland, België en Duitsland. Bij het integratieproces van deze vestigingen is het een groot voordeel geweest dat de data op één centrale plaats in het data warehouse stonden. De data koppelen en afstemmen op elkaar heeft relatief weinig tijd in beslag genomen. SPSS wordt ingezet op zowel macro- als microniveau van het inkoopbeleid. Johan Vlot, Inkoop Manager: "SPSS wordt gezien als een belangrijke tool met dé omzet- en inkoopwaardeinformatie. Dit geeft al aan dat de data en de BI tool betrouwbaar zijn."

Optimaliseren e-Commerce activiteiten en maximaliseren online-omzet

Een belangrijke doelstelling voor Ahrend is het verhogen van de marge. De afdeling e-Commerce & Marketing draagt hiertoe bij door het percentage online verkopen te verhogen. Bart Willem Nagel, e-Commerce Productmanager, meet het bestelgedrag van klanten die zowel via internet als via de normale weg bestellen. Nagel: "Alles is aan te passen in de 'views', zonder dat er een applicatiebeheerder aan te pas hoeft te komen. Om mijn maandelijkse rapportage te maken, heb ik geen dag meer nodig, maar binnen een uur ben ik klaar." De informatie die naar voren komt in deze rapportages is zowel op individueel- als groepsniveau.

Kort samengevat worden de diverse afdelingen door de SPSS oplossing snel en gemakkelijk voorzien van de benodigde (management) informatie. OLAP kubussen helpen bij het monitoren van activiteiten en maken werknemers attent op de nog te ondernemen activiteiten. Door een adequate informatievoorziening treedt Ahrend Office Products International slagvaardig op. Dit heeft zeker bijgedragen aan een stijging van het bedrijfsresultaat van meer dan 30% in 2002 in tijden van laag conjunctuur.

Interesse?

Wilt u meer weten over de oplossing bij Ahrend? Fax dan het antwoordformulier retour en u ontvangt vrijblijvend meer informatie.

"Ahrend beschikt met ShowCase Suite from SPSS letterlijk en figuurlijk over up-to-date 'klantspecifieke informatie'. Veel van de informatie betreft het koopgedrag van klanten van Ahrend, waarbij het inkoopproces bij de klant wordt geoptimaliseerd. Tegelijkertijd is Ahrend met deze oplossing ook in staat de verschillende interne gebruikers van op maat gesneden informatie te voorzien."

Pieter Schotte, Controller Office Products Europe

Wat kunt u verwachten van de nieuwste versie van SPSS: 13.0

SPSS 13.0 ... de nieuwste release van SPSS. Hiermee kunt u nog meer dan versie 12.0 u al te bieden had. De belangrijkste zaken zijn: verbeterde rapportages, makkelijker identificeren van groepen, meer predictive analytics en meer mogelijkheden in data & output management.

Verbeterde rapportages

De grafische engine, die nieuw was in de vorige versie, is verbeterd waardoor de rapportages er beter uitzien. Er zijn hele nieuwe grafische typen toegevoegd en wat betreft het weergeven van die typen zijn de functionaliteiten uitgebreid - waardoor een grotere flexibiliteit ontstaat. Verder kunt u grafieken, tabellen en kaarten direct exporteren naar MS PowerPoint. Deze veranderingen zijn doorgevoerd in de basismodule. Maar niet alleen de basismodule profiteert van de verbeteringen in de grafische engine, ook SPSS Tables geeft u meer mogelijkheden.



Eén van de nieuwe grafische mogelijkheden in SPSS 13.0

Meer predictive analytics

In SPSS Complex Samples zijn een aantal voorspellende algoritmen toegevoegd. Met deze module kunt u nu dus nog beter uw steekproef samenstellen: u neemt geen random steekproef, maar kijkt naar de samenstelling van de populatie. De toegevoegde algoritmen zijn: GLM voor regressie, ANOVA, ANCOVA en logistische regressie voor binaire en multinominale uitkomsten.

Meer mogelijkheden in data management

SPSS Base 13.0 bevat een aantal nieuwe zaken die zorgen voor een tijdsbesparing en verbeterde accuraatheid. De verbetering is o.a. te vinden bij de open vragen, want bij de respons op deze vragen kunnen nu meer tekens ingevuld worden. Dat is te danken aan de langere 'tekst strings'. Verder is er een Time/ Date Wizard toegevoegd wat tot gevolg heeft dat het gemakkelijker werken is met tijden en data. U kunt middels deze Wizard bijvoorbeeld met een ingangsdatum van een bepaald product berekenen hoeveel maanden organisaties al werken met dit product. Ook kunt u gemakkelijk de leeftijd van respondenten berekenen n.a.v. hun geboortedatum. Met autorecode kunt u een string variabele direct omzetten naar een numerieke. In SPSS kunnen de automatisch aangemaakte categorie codes bewaard worden. Dit betekent dat bij veranderingen in de categorieën, de codes zo veel mogelijk gehandhaafd blijven. Bij het aggregeren van bestanden kunt u nu het resultaat direct in het databestand wegschrijven. Verder is er een nieuwe optie 'Split File'. Deze optie geeft de mogelijkheid om een databestand op rij- of kolomniveau op te splitsen. Handig als u bijvoorbeeld veel variabelen heeft en data gaat manipuleren, waarbij nieuwe variabelen gecreëerd worden aan het einde van het bestand.

Nieuwe module: SPSS Classification Trees

Deze aanvullende module is geheel geïntegreerd in de basismodule en geeft u de mogelijkheid om groepen te identificeren, relaties tussen groepen te ontdekken en uiteindelijk voorspellingen over



Beslisboom in SPSS Classification Trees

gedragingen van groepen te doen. De beslisbomen zijn heel visueel, wat tot gevolg heeft dat de resultaten voor iedereen direct duidelijk zijn - ook voor niet-technische mensen die alleen het resultaat willen zien. SPSS Classification Trees bevat vier verschillende algoritmen: CHAID, exhaustive CHAID, CRT en QUEST. Middels deze verschillende algoritmen heeft u de mogelijkheid om te kijken welk algoritme het beste aansluit op uw data.

Interesse?

Wilt u meer weten over wat er nieuw is in SPSS 13.0? Of bent u geïnteresseerd in de nieuwe module SPSS Classification Trees? Faxt u dan het antwoordformulier retour voor meer informatie.

tree_e	credit.sav	- SPSS Da	a Editor		. 🗆 🗙
Ele Edit	View D	ata Iransfor	m <u>Analyze</u> <u>G</u> raphs <u>U</u>	tilities Add-ons Window	Help
		001			011
1 : Credit_	ating		0		and the second sec
	NodeID	Predicted Value	PredictedProbability_ 1	PredictedProbability_ 2	v •
1	9.00	1.00	.44	.56	
2	8.00	.00	.81	.19	
3	1.00	.00	.82	.18	
4	1.00	.00	.82	.18	
5	9.00	1.00	.44	.56	
6	9.00	1.00	.44	.56	
7	9.00	1.00	.44	.56	-
	ata V 4				1
			SPSS	Processor is ready	1

Met SPSS Classification Trees worden de uitkomsten direct verwerkt in het databestand

Trainingschedule SPSS training till July 2005

	days	April	May	June	July
Product training (€ 450,- a day)					
Introduction to SPSS	2	7-8	11-12	13-14	11-12
SPSS Advanced, Further Data Manipulation	2	18-19			
Update to SPSS 13	1		20		
Presenting Data With SPSS Tables, Introduction	1	14			19
Presenting Data With SPSS Tables, Advanced	1			21	
Clementine	1	4			
Introduction to Clementine	2		2-3		7-8
Programming with SPSS Scripts	1			16	
Analytical training (€ 450,- a day)					
Using Catagorical Data	1				1
Introduction to Statistics	2	25-26		23-24	25-26
Variation analysis with SPSS	2				27-28
Advanced Techniques: Regression	2		24-25		
Introduction to Answer Tree	1		31		
Workshops (€ 1175,-)					
Data Mining (Crisp-DM)	2			2-3	

TRAININGEN

Leer de basisbeginselen van SPSS middels de training 'Introduction to SPSS'

Wilt u starten met het gebruik van SPSS? Of heeft u nog weinig ervaring in het werken met de software? En wilt u in 2 dagen efficiënt en productief met SPSS leren werken? Dan kunnen wij u de training 'Introduction to SPSS' aanbevelen.

Deze training wordt in het Vlaams gegeven en vindt plaats op 7-8 april, 11-12 mei, 13-14 juni en 11-12 juli a.s.

Wilt u de meer geavanceerde mogelijkheden van SPSS leren?

Heeft u al enige kennis van SPSS, maar wilt u deze verder uitdiepen? Dan bevelen we u de training 'SPSS Advanced, Further Data Manipulation' aan. Deze training beslaat 2 dagen.

U volgt deze training in het Vlaams op 18-19 april a.s.

Inschrijven?

Wilt u zich inschrijven voor trainingen? Faxt u dan het antwoordformulier aan ons retour.

Na inschrijving ontvangt u een bevestiging en routebeschrijving.