



Potchefstroomse Universiteit.  
vir Christelike Hoër Onderwys

---

WETENSKAPLIKE BYDRAES  
REEKS H: INOUGURELE REDE-NR. 151

**BEDRYFSWISKUNDE EN INFORMATIKA:  
KENNISBLOOTLEGGING EN ANDER UITDAGINGS  
IN DIE 21STE EEU**

**Prof. PJ de Jongh**

Publikasiebeheer Komitee  
Potchefstroomse Universiteit vir Christelike Hoër Onderwys  
POTCHEFSTROOM  
2520

Die Universiteit is nie vir menings in hierdie publikasie aanspreeklik nie.

Navrae in verband met *Wetenskaplike Bydraes* moet gerig word aan:

Die Publikasiebeheerkomitee  
Potchefstroomse Universiteit vir Christelike Hoër Onderwys  
POTCHEFSTROOM  
2520

Kopiereg © 1998 PU vi

**ISBN 1-86822-339-6**

# **Bedryfswiskunde en Informatika: Kennisblootlegging en ander uitdagings in die 21ste eeu**

deur

**PJ de Jongh**

## **1. Inleiding**

Die doel van 'n inougurele rede is, soos ek dit verstaan, om die gehoor op hoogte te bring van die vakgebied waarin ek aangestel is en om op 'n verstaanbare wyse die essensie van die vakgebied en die ontwikkelinge in en betreffende die vakgebied oor te dra. Ek is aangestel as professor in Bedryfswiskunde en Informatika (BWI) in die Sentrum vir Bedryfswiskunde en Informatika. Die Sentrum vir Bedryfswiskunde en Informatika is onlangs by die Potchefstroomse Universiteit vir CHO gestig met as hoofdoel die bedryfsgerigte opleiding van studente in die wiskundige, rekenaar- en ekonomiese wetenskappe. 'n Belangrike motivering vir die stigting van die sentrum was om veral vir studente wat 'n voorliefde vir die wiskundige wetenskappe het, 'n professionele loopbaan in die bedryf te bied. BWI studente word dus opgelei in verskillende aspekte van vakgebiede soos Wiskunde, Toegepaste Wiskunde, Statistiek, Rekenaarwetenskap, Ekonomie en Bedryfsekonomie. Die fokus van die opleiding val op die integrering van hierdie vakgebiede ten einde te verseker dat afgestudeerde studente 'n professionele loopbaan kan volg. Dit is duidelik dat Bedryfswiskunde en Informatika 'n geweldig wye veld dek, en daarom is dit vir my onmoontlik om selfs net al die belangrikste ontwikkelinge en uitdagings aangaande die gebied te probeer gee. In die lig van bogenoemde het ek besluit om hierdie rede soos volg in te deel. Eerstens gaan ek 'n kort oorsig en agtergrond gee van die ontstaan en stigting van die Sentrum van Bedryfswiskunde en Informatika. In die oorsig sal ek die vakgebied Bedryfswiskunde en Informatika en 'n loopbaan as bedryfswiskundige of bedryfsinformatikus bespreek. Die bespreking sal toegelig word deur my eie ervaring op hierdie gebied. Hierna, sal kwessies, geleenthede en uitdagings wat vir Bedryfswiskunde en Informatika voorlê, bespreek word. Laastens sal ek konsentreer op een van die belangrikste geleenthede en uitdagings wat vir die bedryfswiskundige voorlê, naamlik die blootlegging van kennis in groot databasisse.

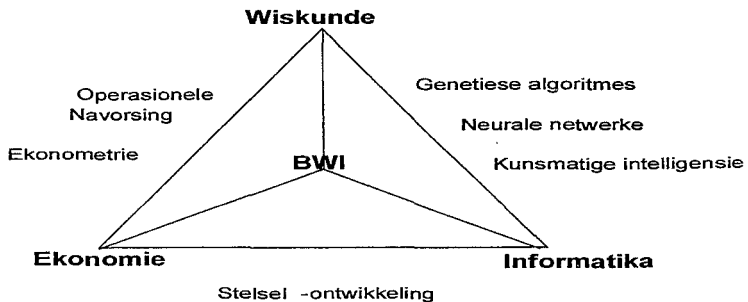
## 2. Bedryfswiskunde en Informatika: Oorsig en agtergrond

In hierdie gedeelte gee ek 'n kort oorsig oor Bedryfswiskunde en Informatika. Sover my kennis strek is die Vrije Universiteit in Amsterdam die eerste universiteit wat 'n kursus in Bedryfswiskunde en Informatika aanbied. Die suksesse wat daar behaal is, met die idee van geïntegreerde en bedryfsgerigte opleiding in die vakgebied, was van die belangrikste redes waarom 'n soortgelyke kursus by Potchefstroom ingestel is. Een groot sukses, voortvloeiend uit die instelling van die kursus aan die Vrije Universiteit, was die toename in studentegetalle in die wiskundige wetenskappe. Die idees betreffende Bedryfswiskunde en Informatika wat in die volgende afdeling bespreek word, is grotendeels afkomstig van die model wat aan die Vrije Universiteit gevolg word.

### 2.1. Bedryfswiskunde en Informatika as vak

In sy inaugurele rede by die Vrije Universiteit in Amsterdam gee Kersten (1993) sy gedagtes oor die posisionering van Bedryfswiskunde en Informatika. Hieronder volg 'n opsomming van die hoofgedagtes soos deur hom uitgespreek.

Kersten beskou Bedryfswiskunde as een komponent van die vakgebied Bedryfswiskunde en Bedryfsinformatika. Volgens hom het Bedryfswiskunde te doen met die gesamentlike gebruik van wiskundige en rekenaarwetenskaptegnieke en teorie om vraagstukke wat in die bedryfslewe voorkom, op te los. In Figuur 1 stel hy een weergawe voor van die posisionering van BWI.



**Figuur 1:** Die BWI-driehoek

Die hoekpunte van die driehoek word gevorm deur die vakgebiede Wiskunde (ook Toegepaste Wiskunde en Statistiek), Rekenaarwetenskap en Ekonomie (ook Bedryfseconomie). In die swaartepunt van die driehoek is die vakgebied

Bedryfswiskunde en Informatika. Op die as wat die vak Wiskunde met die vak Ekonomie verbind plaas Kersten 'n vakke soos Operasionele Navorsing en Ekonometrie wat, volgens hom, min rekenaarwetenskap-inhoud het. Op die as wat Rekenaarwetenskap met Ekonomie verbind plaas hy die vak Bestuursinformatika, wat geen wiskundige inhoud bevat nie. Laastens, op die as wat Rekenaarwetenskap met Wiskunde verbind plaas hy vakke soos Kunsmatige Intelligensie, Neurale Netwerke en Genetiese Algoritmes. Die vak Bedryfswiskunde vorm nou 'n ewewig tussen die vakke Wiskunde, Ekonomie en Rekenaarwetenskap en integreer die vakke met spesifieke fokus op die oplos van bedryfsprobleme.

Soos wat die vak hierbo beskryf word, kan dit myns insiens nouliks as 'n aparte volwaardige vak gemotiveer word. Dit verskil, wat my betref, weinig van wat die aanvanklike oogmerk met Operasionele Navorsing was en nog steeds is.

## **2.2. Bedryfswiskunde en Informatika as beroep**

Die vak Bedryfswiskunde en Informatika, soos voorgestel in Figuur 1, laat onwillekeurig vrae opduik: Wat is die beroep van iemand wat in die vak opgelei is? Is dit 'n bedryfswiskundige/informatikus, 'n bedryfswiskundige of 'n bedryfsinformatikus of wat? Wat doen so iemand eintlik? Wat maak hierdie beroep anders as ander beroepe wat raakvlakke met hierdie vakke het? Byvoorbeeld, beroepe soos die van bedryfsingenieur, operasionele navorser, statistikus, ekonometris, ekonoom, rekenaarwetenskaplike en so meer. Alhoewel Kersten nie hierdie vrae behandel in sy inougorele rede nie, is daar myns insiens nie eenvoudige antwoorde nie. As 'n bedryfswiskundige breedweg gesproke wiskundig en rekenaar gebaseerde tegnieke en teorie gebruik om bedryfsprobleme op te los, maak dit die bedryfswiskundige weinig verskillend van 'n bedryfsingenieur, operasionele navorser en selfs 'n statistikus. Die bedryfsingenieur het spesifiek ten doel die verbetering van die doelmatigheid en doeltreffendheid van ondernemings en word in die vakke wat in die BWI driehoek verskyn, geïntegreerd en bedryfsgerig opgelei. Omdat die operasionele navorser tipies dieselfde oogmerk het, was en is die beroep gedurig in 'n identiteitskrisis oor wat 'n operasionele navorser nou eintlik is en moet doen en hoe dit die operasionele navorser van ander beroepe onderskei. Operasionele navorsingskonferensies deur die jare is gekenmerk deur bespreking van hierdie beroepskwessies. Die titel van 'n artikel wat 'n toonaangewende operasionele navorser reeds jare gelede geskryf het, illustreer hierdie dilemma op 'n uiterste wyse. Kyk Ackoff (1979): *The Future of*

*Operations Research is Past.* Daar was egter ook suksesse met die opleiding van operasionele navorsers. Volgens Gass (1996) is die enigste suksesvolle program, wat teorie en praktyk suksesvol integreer, dié van Woolsey by die Colorado School of Mines (kyk Woolsey en Maurer 1995). Die program het dit reggekry om bedryfswiskundiges (of dan operasionele navorsers) op te lei wat genoeg teoretiese kundigheid en algemene probleemoplossingsvaardighede het om komplekse bedryfsprobleme suksesvol op te los.

Die beroep van die statistikus is duideliker, maar eng efinieer rakende die vak. Statistiek, en hoewel statistici tipies nie identiteitskrisse ervaar nie (dit is dié wat as statistici werk), was daar in die verlede nie 'n oorloed werksmoontlikhede vir statistici nie. Dit word verder aangehelp deur die beskikbaarheid van gesofistikeerde sagteware wat maklik deur bedryfsingenieurs en ander beroepslui gebruik kan word en dus 'n bedreiging vir die beroep inhou. In De Jongh en De Wet (1993) en meer onlangs in die presidensiële rede van Kettenring (1997) van die Amerikaanse Statistiese Vereniging word spesifiek daarop gewys dat die vak Statistiek kerngesond is, maar dat die beroep as statistikus bedreig word as statistici nie iets daadwerkliks daaraan doen nie. Voorstelle word in beide artikels gemaak oor wat statistici te doen staan.

Dus, in die lig van bostaande kan ons maklik in die slagkat trap om 'n bedryfswiskundige op te lei wat direk met goed gevestigde professionele beroepe soos die bedryfsingenieur en rekenaarwetenskaplike (stelselanalisis) kompeteer en uiteindelik met dieselfde probleme as dié van operasionele navorsers en statistikus eindig. So 'n poging is nie die moeite werd nie, maar wat is die oplossing? Myns insiens moet die bedryfswiskundige opgelei word om 'n nismark te bedien. In so 'n mark moet daar relatief min kompetisie van ander beroepe bestaan, en die bedryfswiskundige moet gespesialiseerde kundigheid in Wiskunde, Ekonomie en Rekenaarwetenskap hê, maar met spesialisasie in 'n bepaalde toepassingsgebied ten einde probleme in dié area vinnig en doeltreffend op te los. *Fokus op 'n bepaalde industrie en toepassingsgebied is van die allergrootste belang wanneer bedryfswiskundiges opgelei word.* Daar kan in dié sin ook met vrug na die opleidingsprogram van die Colorado School of Mines gekyk word.

### **2.3. Bedryfswiskunde en Informatika: My ervaring**

Ek is self 'n tipiese voorbeeld van iemand wat in die wiskundige, ekonomiese en rekenaarwetenskappe opgelei is en daarna 20 jaar se ondervinding in die semi-staat en privaatsektor opgedoen het. In hierdie tyd het ek waardevolle

ondervinding opgedoen as konsultant, projekbestuurder en bestuurder. Ek wil hierdie ervaring graag kortliks met u deel en dit gebruik om my siening van Bedryfswiskunde en Informatika as vak en beroep te verduidelik.

My eie loopbaan het as statistikus begin en het aanvanklik gehandel oor die lewering van 'n statistiese konsultasiediens aan ingenieurs. Alhoewel ek toegepaste navorsing in Statistiek gedoen het, was my eintlike belangstelling in die konsultasietake waar Statistiek 'n belangrike bydrae in die oplos van ingenieursprobleme gespeel het. Dikwels het die werk wat ek en my kollegas gedoen het, daartoe bygedra dat beter besluite in belangrike projekte geneem kon word. Ek het egter vroeg in my loopbaan, en nogal met 'n skok, agtergekom dat Statistiek ( en dus ook Wiskunde en Toegepaste Wiskunde) 'n baie klein rol in die oplos van die meeste werklike probleme speel. Die grootste deel van dié probleme het dikwels die strukturering van die probleem en die verstaan van die toepassingsgebied behels. Die klein rol wat Statistiek in die oplos van werklike probleme speel, het daartoe gelei dat ek 'n mate van frustrasie ervaar het oor my rol as statistikus wat ek as baie beperkend beskou het, veral ten opsigte van loopbaanmoontlikhede in die organisasie. Ek wou 'n baie groter rol speel in projekte of take wat probleme aanpak wat vir die bestuur van maatskappye saak maak. In daardie stadium het operasionele navorsing gewild begin word as 'n vak wat die wiskundige wetenskappe en Rekenaarwetenskap geïntegreerd aanwend om probleme op 'n multidissiplinêre wyse aan te pak en op te los.

As operasionele navorser het ek nou binne multi-dissiplinêre spanne saamgewerk om probleme op te los. Gou het die belangrikheid van die sagter wetenskappe in algemene probleemoplossing my opgeval, byvoorbeeld projekbestuur, strategiese beplanning, prosesverbetering, mensbestuur en veral fasiliteringsvaardighede. Die snel ontwikkeling in rekenaartegnologie het ook meegebring dat meer geleenthede ontstaan het, geleenthede om oplossings vir probleme in produkte (rekenaarstelsels) te verpak. As operasionele navorser by *die Instituut vir Maritieme Tegnologie* het ek betrokke geraak in die ontwikkeling van besluitsteunstelsels. Ek het opgetree as projekteier of projekbestuurder van sodanige projekte en in die proses intens bewus geword van die belangrikheid van rekenaartegnologieë soos databasisbestuur, databasisontwerp, stelselontwikkelingsmetodologieë en kunsmatige intelligensie (byvoorbeeld kennisstelsels en neurale netwerke). Ek kon nou hierdie kennis gebruik om wiskundige modelle en tegnieke in die rekenaar te implementeer. Dit het vir my en ook vir die kliënt baie bevrediging verskaf, want nou kon ek tasbare produkte vir

kliënte aflewer wat hulle in staat stel om beter besluite te neem. Kyk byvoorbeeld De Jongh et al.(1994) in hierdie verband. Hierdie stelsels was aanvanklik geskoei op besluitsteun op operasionele vlak, maar later was ek ook betrokke as bestuurskonsultant in die ontwikkeling van besluitsteunstelsels om strategiese besluite te steun. By *Deloitte and Touche Consulting Group* was ek as bestuurskonsultant betrokke by 'n operasionele navorsingspan wat vir seker een van die suksesvolste en definitief grootste operasionele navorsingsprojek nog in Suid-Afrika verantwoordelik was. Hierdie projek, genoem OPTIMUM, het die Franz Edelman medalje, 'n toekenning wat jaarliks deur die *Institute for Operational and the Management Sciences (INFORMS)* in die VSA vir die beste bydrae in die bestuurswetenskappe gemaak word, gewen. Kyk et al. en Wessels (1997). Dit is interessant om daarop te let dat Wessels, wat projekteier van hierdie projek was, 'n produk van die operasionele navorsingsprogram van die *Colorado School of Mines* is.

Gedurende die ontwikkeling van my loopbaan het ek in veral die beginstadium een of twee keer 'n identiteitskrisis beleef. Byvoorbeeld: Wat is ek nou eintlik, 'n statistikus, 'n operasionele navorser, 'n stelselontwikkelaar of 'n bestuurskonsultant? Later in my loopbaan het hierdie vrae my al minder gepla totdat ek tot die besef gekom het dat dit eintlik nie regtig saak maak nie. Bestuur stel belang in 'n individu wat pro-aktief bestuursprobleme kan oplos deur van die nuutste en geskikste tegnologie gebruik te maak. *Sodanige individu moet alltyd bereid wees om nuwe kennis en ervaring op te doen binne en buite die veld wat aanvanklik gestudeer is. Wat uiteindelik van belang is, is wat jy kan doen en nie noodwendig wat jy gestudeer het nie.* Dit is egter so dat diegene wat in die ekonomiese, wiskundige en rekenaarwetenskappe gestudeer het, 'n baie goeie fondament het om vir maatskappye tot nut te wees en die vermoë het om gou aan te pas by veranderende omstandighede. Ook is dit goed dat die individu 'n akademiese tuiste het by en tuis voel in 'n professionele vereniging van 'n vak waarin hy opgelei is.

### **3. Bedryfswiskunde en Informatika: Kwessies en uitdagings**

Voordat uitdagings vir Bedryfswiskunde en Informatika bespreek word, behoort eers gekyk te word na veranderinge wat in die omgewing (waarbinne ons optree) plaasvind, die behoeftes wat as gevolg daarvan ontstaan en hoe Bedryfswiskunde en Informatika daardeur geraak word.



### 3.1. Invloed van die omgewing

Tans (einde 20ste eeu) vind snelle ontwikkelinge plaas op die gebied van rekenaartegnologie en kommunikasietegnologie. Die koms van die Internet en die selfoon het veroorsaak dat enigiemand op bykans enige plek met iemand anders kan kommunikeer.

Veral die internet beïnvloed besigheid en die doen van besigheid dramaties. Met behulp van 'n selfoon en 'n draagbare rekenaar kan 'n persoon sake doen met enige maatskappy wat op die Internet gekoppel is. Alreeds kan banksake, boek- en kompakskyfaankope en vlugbesprekings op die Internet gedoen word. Nuwe besighede ontstaan daagliks op die Internet wat met innoverende maniere kom om sake te doen. Oorsese universiteite en instansies bied kursusse aan wat op die Internet geneem kan word. Dit alles lei tot sogenaamde globalisering en verhoogde mededinging. Die mark vir enige produk is enigiemand in die wêreld, en die kompetisie vir die verskaffing van enige produk of diens is enigiemand in die wêreld wat 'n soortgelyke produk of diens verskaf.

Dus, hoewel ons in Suid-Afrika skielik 'n baie groter mark vir ons produkte en dienste het, het ons ook baie strawwer mededinging. Vir ontwikkelende ekonomieë (soos Suid-Afrika) het verhoogde mededinging ernstige implikasies, omdat daar met ontwikkelde ekonomieë meeding moet word. Trompop mededinging kan katastrofies wees, maar gelukkig bied alliansievorming 'n alternatief.

Wêreldwyd is daar 'n groot vraag na hooggeskoolde arbeid; veral in die rekenaar-omgewing word graduandi opgeraap. Ook is daar skielik weer 'n oplewing in die behoefte aan opgeleide statistici wat veral in die finansiële industrie benodig word om te help met risikobestuur. Tydskrifte soos Finansiële Tegniek, Financial Mail en andere bevat advertensies vir beroepe in bogenoemde areas (dikwels met gepaardgaande astronomiese salarisse). Ook is hierdie advertensies dikwels vir poste in internasionale maatskappye. Dus, as deel van globalisering kan enigiemand met die regte opleiding baie makliker oorsee werk kry en is daar ook wat afgestudeerde studente betref, 'n baie groter mark as die tradisionele. In Suid-Afrika begin ons die uitwerking van globalisering voel en het ons ook ander nasionale vereistes waaraan aandag gegee moet word, soos dat maatskappye verteenwoordigend van die mense van Suid-Afrika moet wees. In die lig van bogenoemde is daar vandag baie meer druk op maatskappye om mededingend te wees en te oorleef in 'n al hoe mededingender mark.

### 3.2. Behoeftes van die bedryf

Die winsmotief en oorlewingsdrang bly van die belangrikste dryfkragte agter enige maatskappy in die privaatsektor. Maatskappye wil dus winste maksimeer, en dit kan gedoen word deur inkomste te maksimeer en uitgawes te minimeer. Inkomste kan vermeerder word deur byvoorbeeld nuwe produkte of dienste te verkoop en meer bestaande produkte of dienste te verkoop. Uitgawes kan verminder word deur bestaande produkte of dienste op 'n koste-doeltreffender wyse te lewer. Bogenoemde is die stukrag agter die meeste ander probleme waarmee maatskappye in die verlede geworstel het, nou worstel en in die toekoms sal worstel. Bestuurskonsultasiemaatskappye kom dikwels met nuwe probleemoplossings metodologieë vorendag wat daarop gemik is om onderhawige probleme op te los. Van die mees onlangse is strategiese beplanning (Engels: *Strategic planning*), totale-gehalte-bestuur (Engels: *Total Quality Management*), besigheidsprosesherontwerp (Engels: *Business Process Re-engineering*), datapakhuis (Engels: *Data Warehousing*) en kennisblootlegging in databasisse (Engels: *Knowledge Discovery in Databases*), om maar 'n paar te noem. Tans is daar 'n groot mark vir transformasieprojekte in Suid-Afrika wat 'n sogenaamde totale oplossing tot gevolg het: Nuut herontwerpte besigheidsprosesse, ondersteun deur rekenaarsistels wat die regte bestuursinligting op die regte tyd en op die regte plek vir die regte besluitnemer (regte vlak) beskikbaar stel, sodat die hele organisasie op die doeltreffendste wyse kan fungeer. Miljoene rande word geïnvesteer in pogings om wins te maksimeer en uiteindelik inkomste te vermeerder en/of uitgawes te verminder.

#### **Implikasies vir Bedryfswiskunde en Informatika**

Wat is die bydrae wat Bedryfswiskunde en Informatika kan maak om bogenoemde probleme aan te pak? Daar kan ten opsigte van baie aspekte bydraes gelewer word, en ek noem enkele:

- Statistiese en finansiële wiskundige tegnieke kan gebruik word om maatskappye te help om risikobestuur te verbeter. Dit is om mark-, krediet-, operasionele, likiditeit- en versekeringsrisiko's te minimeer.
- Statistiese en kunsmatige intelligensietegnieke kan gebruik word vir kennisblootlegging in groot databasisse.

- Optimeringstegnieke kan in strategiese beplanning gebruik word om maatskappye te help om waarde vir geld te kry, met ander woorde die beste uitset, gegee die beskikbare hulpbronne, te verkry.
- Simulasie- en modelleringstegnieke kan gebruik word om besigheidsprosesse te ondersoek en te verbeter.
- Besluitsteunstelsels kan ontwikkeling word om besluitneming in operasionele en strategiese prosesse te ondersteun.
- Statistiese en ekonometriese tegnieke kan gebruik word vir die evaluering van beleggings en om investeringsbesuite te steun.

In gesprekke met rolspelers in die finansiële industrie is dit duidelik dat die grootste behoefte tans kundigheid rakende risikobestuur is. 'n Ander verwante en belangrike, maar sekondêre, behoefte is die blootlegging van kennis in groot databasisse. Hierdie twee behoeftes moet myns insiens die fokuspunt van opleiding in die Bedryfswiskunde- en Informatikaprogram wees, maar egter nie ten koste van die basiese dissiplines nie.

### **Implikasies vir 'n loopbaan as bedryfswiskundige**

In die lig van die verandering in die omgewing waarbinne maatskappye optree, wat is die verwagtinge wat maatskappye deesdae en in die toekoms van werknemers koester en spesifiek die bedryfswiskundige?

Volgens Dalton & Thomson (1986) is daar 4 stadia in die loopbane van professionele beroepe, nl.

- vakleerling (iemand wat in opleiding is),
- kollega (iemand wat onafhanklike bydraes lewer),
- mentor (iemand wat ander bestuur en mense ontwikkel) en
- strateeg (iemand wat die onderneming rig).

Persone met tegniese opleiding betreffende 'n bepaalde vakgebied soos Statistiek, Toegepaste Wiskunde, Rekenaarwetenskap en Ekonomie kan as probleemoplossers binne 'n bepaalde vakgebied beskou word. Probleemoplossers beweeg selde bokant die kollegastadium. Dié wat dit wel doen, vul tegniese kennis aan met bestuurskennis en bestuursvaardighede en fokus meer op die "groter" kwessies. Hoe verder die werknemer vorder in

die rigting van strateeg, hoe groter is die uitwerking van wat hy doen op die organisasie. 'n Verdere interessante waarneming is dat gedurende die ontwikkeling van die individu in sy beroep, dit al hoe belangriker word om met mense saam te werk, te netwerk, ens. Dit is so omdat die kompleksiteit van die probleme sodanig is dat een mens alleen dit nie kan oplos nie. In ooreenstemming hiermee is die verskuiwing van individuele probleemoplossing of navorsing na projekte waar spanne saamwerk om probleme op te los. Hoe groter die projek, hoe groter die span en die moontlike uitwerking op die organisasie. *Opleiding in die natuurwetenskappe en spesifiek die vakrigtings waarop Bedryfswiskunde en Informatika gebaseer is, fokus tipies op probleem-oplossing, is eng en na binne gefokus en het om daardie rede verstaanbaar min invloed of uitwerking op die besigheid of onderneming. Universiteite, en in dié geval die Sentrum vir BWI, kan baie doen om hierdie persepsie te verander.* Dit is ook belangrik om daarop te let dat hoe verder die individu in sy loopbaan vorder, hoe minder belangrik word wat hy gestudeer het, maar hoe belangriker wat hy weet, hoe hy dit aanwend en hoe die onderneming daardeur baatvind. 'n Direkteur van 'n onderneming is in die eerste plek nie 'n statistikus, toegepaste wiskundige, ingenieur of rekenmeester nie, maar die direkteur van die onderneming.

Hieronder volg 'n paar vereistes wat besighede aan werknemers stel bo en behalwe die kundigheid in die veld waarin hulle opgelei is. Hahn en Hoerl (1998) stel die volgende baie belangrike verwagting van bestuur, en ek haal aan: *Business is no longer willing to have isolated specialists on their pay-rolls. Those that remain need to apply their time to customer projects for specific tasks performed. The key aim is to be responsive and vital to today's business needs.* In kommentaar op dieselfde artikel noem Snee die volgende verdere vereistes:

- Die tydigse lewering van resultate is van groot belang. Dit is beter om goeie resultate betyds te lewer as die beste (optimale) resultate te laat.
- Die werknemer en konsultant moet die besigheidsumgewing waarin die maatskappy optree, ken en verstaan. Dit veronderstel die verstaan van die kliënte, markte, produkte en rolspelers.
- Die werknemer moet projekspanne kan lei en fasiliteer.

- Kennis oor groepdinamika en die bestuur van interpersoonlike verhoudinge is van groot belang.
- Werknemers moet die besigheidsprosesse verstaan en metodes ken om prosesse te verbeter.
- Kennis van hoe projekte bestuur, beplan en uitgevoer moet word, is van groot belang.
- Die ontwerp en bestuur van vergaderings is noodsaaklik.
- Werknemers wat goed kan fasiliteer en die vermoë het om gedagtes en idees te struktureer, is van groot waarde vir ondernemings.
- Laastens is dit belangrik dat werknemers kreatief dink, 'pro-aktief' optree en graag in die betekenisvolle en belangrike projekte betrokke raak.

In sy presidensiële rede as president van die Amerikaanse Statistiese Vereniging, voeg Kettenring(1997) by dat statistici (en dus wiskundiges en andere) hul bevindings op 'n maklik verstaanbare wyse aan kliënte moet kan oordra.

In die *Mail & Guardian* van 21 tot 27 Augustus word verskillende eienskappe van die graduandi wat besighede soek, opgesom. Die eienskappe wat genoem is, en ek haal aan, is:

- *Businesses need people who are multi-skilled and can be fully employed by performing different tasks. Graduates will have to work consistently well and should know that they will only be as good as their last project.*
- *Businesses recruit people who are eager to learn and who keep abreast of the latest technologies and trends which they can bring to the job and which will give their business the cutting edge over competitors.*
- *The more flexible and productive a person the greater the chances of employment.*

Myns insiens moet die belangrikheid van besigheidsetiek ook beklemtoon word en moet studente die basiese beginsels van gesonde besigheidsetiek en Christelike norme geleer word. Veral agtergeblewe gemeenskappe wat nie die voorreg van 'n Christelike opvoeding gehad het nie, kan baie baat vind daardeur. Hierin kan die

Potchefstroomse Universiteit vir CHO 'n groot rol speel wat vir ons land op die lang termyn van groot waarde kan wees.

### **3.3. Samevatting van die uitdagings vir Bedryfswiskunde en Informatika**

In die voorafgaande bespreking het ek genoem dat ek nie Bedryfswiskunde en Informatika as 'n vakgebied in eie reg beskou nie. Intendeel, dit is baie nou verwant aan Operasionele Navorsing (*Engels: Operational Research / Management Science*), want beide is gebaseer op die onderliggende vakke Ekonomie, Wiskunde en Rekenaarwetenskap wat gebruik word om probleme in die bedryf op te los. Die beroep bedryfswiskundige is ook nie uniek nie, en bedryfsingenieurs, operasionele navorsers, statistici, rekenaarwetenskaplikes en ekonome sou wou glo dat baie van die probleme wat in die bedryf opduik, net so goed deur hulle gehanteer sou kon word. Wat nodig is, is fokus in opleiding van die bedryfswiskundige om meer gespesialiseerde markbehoefte aan te pak. Daar moet op 'n bepaalde industrie gefokus word, en kernprobleme in daardie industrie moet gebruik word om opleiding te rig. Omdat omstandighede verander, sal markbehoefte verander, en dus sal opleiding ook moet aanpas.

Tans is ons fokus met reg op die finansiële industrie en om behoeftes daar aan te pak. Die grootste behoefte tans is die opleiding van sogenaamde finansiële ingenieurs om finansiële instellings met die bestuur van mark-, krediet-, operasionele en versekeringsrisiko behulpsaam te wees. In dié verband is dit ook nodig om sogenaamde datageoloë op te lei wat kennis, verskuil in massiewe databasisse, kan blootlê sodat maatskappye dit kan gebruik om bostaande risiko's te minimeer. (Kennisblootlegging en die uitdagings in hierdie verband sal in meer besonderhede in afdeling 4 bespreek word.) In die toekoms sal 'ons voortdurend moet monitor of die opleiding nog geldig is en of dit nog steeds die behoefte in die mark bevredig. 'n Aanduiding sou die bereidwilligheid van maatskappye wees om beurse aan studente toe te ken. Myns insiens is die volgende sake voorvereistes vir goeie opleiding as finansiële ingenieur of datageoloog.

*Ten eerste is 'n goeie basiese onderlegtheid in die teoretiese onderbou van Wiskunde, Toegepaste Wiskunde, Statistiek, Ekonomie en die rekenaar ononderhandelbaar. Dan moet studente ook goed onderrig word in die integrasie van die basiese vakke vir oplossing van probleme in die bepaalde toepassingsveld, hetsy risikobestuur of kennisblootlegging. Dit moet aangevul word met belangrike vaardighede wat die bedryf vereis, nl. projekbestuur, vergaderingsprosedures, fasilitering, dokumentasie en veral die aankweek van kreatiewe denke en pro-*

aktiwiteit. 'n Afgestudeerde student met hierdie vaardighede sal 'n aanwinst wees vir enige maatskappy. Goeie opleiding kan 'n verskil maak.

Samewerking met die onderliggende basiese vakke en beroepe wat daaruit voortvloei, is van kardinale belang. 'n Statistikus of rekenaarwetenskaplike met belangstelling in kunsmatige intelligensie kan maklik deel word van hierdie program en so sy ervaringsveld verbreed. Dit is egter van belang dat die kursus in Bedryfswiskunde en Informatika op hierdie onderliggende vakke berus en daarsonder nie 'n eie bestaansreg het nie. Ook is die beroepe datageoloog en in 'n mindere mate die finansiële ingenieur baie nou verwant aan dié van statistikus en operasionele navorser en selfs rekenaarwetenskaplike (spesialisasie kunsmatige intelligensie). Hierdie opleidingsprogram kan dus die tradisionele beroepe van operasionele navorser en statistikus bevoordeel in dié opsig dat meer loopbaanmoontlikhede ontstaan. In die akademiese opset moet ons nie eng na binne in die spesifieke vak konsentreer nie, maar na buite en samewerking oor vakgrense aanmoedig. *Die debat of byvoorbeeld data-ontginning (Engels: Data mining) in Statistiek of Kunsmatige Intelligensie tuishoort, is myns insiens futiel. Gryp eerder die geleentheid vir beide vakke; uiteindelik sal dit tot voordeel wees vir beide beroepe.*

In die volgende ardeing gaan ek een van die grootste uitdagings vir Bedryfswiskunde en Informatika, naamlik kennisblootlegging in databasisse, in meer besonderhede bespreek. Die ander groot uitdaging, naamlik dié van risikobestuur, sal deur die direkteur van die Sentrum vir Bedryfswiskunde en Informatika in sy inaugurele rede gedek word (kyk Erasmus 1999).

#### **4. Kennisblootlegging in groot databasisse**

Kennisblootlegging is 'n snelgroeiende area wat op die vakgebiede Kunsmatige Intelligensie en Statistiek gebaseer is. Tegnieke vir kennisblootlegging is baie wiskundig van aard, en kennis van Wiskunde, Statistiek en Kunsmatige Intelligensie is nodig om hierdie tegnieke prakties toe te pas en ook om nuwe tegnieke te ontwikkel. Kennisblootlegging van groot databasisse kan maatskappye help om 'n wenvoorsprong oor mededingers te bewerkstellig. Dit kan onder meer help dat maatskappye doelmatiger en doeltreffender bemark en risiko doeltreffender bestuur. Byvoorbeeld, kennisblootleggingstegnieke kan help om verdagte transaksies op te spoor ('n voorbeeld van operasionele risiko) en om bestaande en voornemende kliënte se risikoprofiel vas te stel en te ontleed ('n voorbeeld van kredietrisiko).

#### 4.1. Hoekom is kennisblootlegging belangrik?

Reusehoeveelhede data word elke dag deur maatskappye ingesamel. Banke hou boek van elke kliënt se tjek- en kredietkaarttransaksies, telekommunikasiemaatskappye versamel inligting oor die telefoonoproep van kliënte, voedselmaatskappye versamel inligting oor die verkope van voedselsoorte, en finansiële instellings bestudeer die beweging van aandele op effektebeurse. Data word wyd versamel, en individuele transaksies of spesifieke data-elemente word op meer as een databasis geberg. Byvoorbeeld, 'n enkele kredietkaarttransaksie by 'n voedselmaatskappy word geberg op die databasis van beide die voedselmaatskappy en die bank van die individu.

Maatskappye berg hierdie data op reusedatabasisse, genoem datapakhuse. Die datapakhuis is die sentrale databasis van 'n groot maatskappy waar belangrike historiese inligting vasgelê word, hoofsaaklik om die bestuur van maatskappye te help met strategiese en operasionele besluitneming. Aangesien produkte van maatskappye tans baie soortgelyk is (byvoorbeeld banke se produkte is onder andere tjekrekenings, kredietkaarte en verbandenings met min verskille in produkte tussen banke), val die klem tans baie sterk op die lewering van 'n uitstekende diens aan kliënte om sodoende die wenvoorsprong te probeer bekom. Dus dienslewering en goeie kliëntverhoudings word een van die belangrike fokusareas om maatskappye 'n wenvoorsprong bo mededingers te gee. 'n Ander belangrike fokusarea is om deur middel van doelmatige en doeltreffende bemarking 'n groter deel van die mark te bekom.

Maar hoe kan datapakhuse en kennisblootlegging van die data in die datapakhuis maatskappye help om 'n wenvoorsprong op te bou? Kennisblootleggingstegnieke kan maatskappye help om 'n wenvoorsprong te verkry deur antwoorde op die volgende tipe probleem te bekom:

- Die maatskappy wil op die kostedoeltreffendste manier 'n nuwe of bestaande produk bemark. Die vraag wat die maatskappy wil beantwoord, is: Aan wie moet die produk bemark word sodat die maatskappy die hoogste moontlike opbrengs kry? Die vraag kan ook anders gestel word. Wat is die waarskynlikheid dat kliënte (nuwe of bestaande) die bepaalde produk sal koop, en hoe kan die waarskynlikheid gemaksimeer word?



- 'n Finansiële instelling wil graag lenings aan kliënte toestaan maar wil tegelyk sy risiko ten opsigte van slegte skuld beperk. Die vraag wat beantwoord moet word, is: Aan watter kliënte moet lenings toegestaan word? Anders gestel: Wat is die risikoprofiel van 'n kliënt, en hoe moet besluit word oor die toekenning al dan nie van die lening?
- 'n Finansiële instelling wil graag sy kliënte en homself teen bedrog beskerm. Die vraag wat beantwoord moet word is: Watter transaksies kom verdag voor en dui op bedrog? Anders gestel: Wat is die waarskynlikheid dat 'n sekere transaksie dui op bedrog?
- 'n Bank wil graag sy bestaande kliënte behou en wil seker maak dat bestaande kliënte uitstekende diens ontvang. Byvoorbeeld, as 'n kliënt verhuis, kan die oordrag van rekeninge vergemaklik word. Die vraag wat beantwoord moet word, is: Wanneer kan met redelike sekerheid gesê word dat 'n bepaalde kliënt oorweeg om sy rekeninge te sluit?
- 'n Voedselmaatskappy wil sy verkope verhoog. Die vraag wat beantwoord moet word, is: Watter voedselsoorte verkoop tipies saam? Deur dié voedselsoorte saam op winkelrakke te plaas kan dalk lei tot verhoogde verkope.

Dit behoort duidelik te wees dat antwoorde op bostaande vrae die maatskappy beter in staat sal stel om uitstekende gehalte diens te lewer en kostedoeltreffender te bemark, sodat die maatskappy daardeur bestaande (goeie) kliënte sal behou en (goeie) nuwe kliënte kan bykry. Kennisblootlegging poog om bostaande vrae vinnig te beantwoord sodat maatskappye pro-aktief kan optree betreffende bemarking, verkope en dienslewering aan kliënte.

## 4.2. Wat is kennisblootlegging?

Kennisblootlegging kan gedefinieer word as die wetenskap om interessante en nie-ooglopende patrone in massas data te vind. Dit is veral 'n hulp ten opsigte van die bestuur van maatskappye om kliëntbehoefes en verbruikersgedrag beter te verstaan en help die bestuur om pro-aktief op te tree. Soos 'n geoloog na patrone in die aardkors soek, so sal die data-geoloog na patrone in die pakhuis van maatskappye soek en in dié proses antwoorde vind om probleme op te los. Kennisblootlegging het eers onlangs gewild geword as gevolg van die ontwikkeling van rekenaartegnologie wat die mens in staat stel om massas data in 'n oomblik te

verwerk. Dit is, soos voorheen genoem, gebaseer op die vakgebiede kunsmatige intelligensie en Statistiek. Ten einde kennisblootlegging en kennisblootleggingstegnieke beter te verstaan, kyk ons nou in meer besonderhede na dié vakgebiede.

*Kunsmatige intelligensie* is 'n betreklik nuwe vakgebied en het te doen met die strewe van rekenaarwetenskaplikes om intelligente masjiene te ontwikkel wat soos die mens kan leer, dink en op omgewingsimpulse reageer. Gedryf deur die visie om intelligente masjiene of robotte te bou, het rekenaarwetenskaplikes verskeie modelle ontwikkel wat gebaseer is op die manier waarop mense dink en op die biologiese prosesse in die brein. Merkwaardige nuwe tegnieke is ontwikkel soos neurale netwerke (gebaseer op die biologiese werking van die brein), kennisstelsels (gebaseer op die mens se geneigdheid om reëls te gebruik om besluite te neem of afleidings te maak) en outomatiese metodes om reëlinduksie te doen (afleiding van reëls uit data). Suksesse wat tot op hede behaal is, het veroorsaak dat groot verwagtinge geskep is om hierdie tegnieke ook in die besigheidsomgewing toe te pas.

*Statistiek* is 'n welbekende vakgebied wat reeds goed gevestig is. Die basis van Statistiek is die waarskynlikheidsteorie, en dit kan beskryf word as die wetenskap van besluitneming onder 'n toestand van onsekerheid. Statistiek het hoofsaaklik te doen met data-analise en die maak van gevolgtrekkings uit data wat vir navorsers, bestuurders en ingenieurs van belang kan wees. Dit het grootliks ontwikkel in die tyd voor rekenaars, en statistici was in daardie tyd meestal gekonfronteer met relatief klein datastelle. Daar het egter die afgelope paar jaar, aangevuur deur die rekenaar, groot nuwe ontwikkelinge plaasgevind, en nuwe tegnieke wat baie rekenaar afhanklik is, is ontwikkel. Omdat statistici reeds meer as 'n eeu lank besig is met data-analise, is daar ook deur die jare 'n grondige en gevestigde metodologie vir die analise van data ontwikkel. Tegnieke van belang vir kennisblootlegging is trosvorming (groepering van data en veranderlikes in homogene groepe), besluitnemingsbome (klassifikasie van data-elemente in homogene groepe), diskriminantanalise en regressie-analise.

Kunsmatige Intelligensie en Statistiek vorm die hoekstene van kennisblootlegging. Die tegnieke wat uit hierdie vakgebiede spruit, is wiskundig van aard en word prakties met behulp van die rekenaar toegepas. Dus is die datageoloog eintlik 'n toegepaste wiskundige of statistikus wat met behulp van die rekenaar

kennisblootlegging van groot data-basisse doen om patrone te identifiseer of kennis wat opgesluit (versteek) in die data lê, te ontbloot.

Die tipiese probleme wat deur kennisblootleggingstegnieke aangepak word, kan breedweg geklassifiseer word as

- Klassifikasie (bv. klassifiseer 'n leningsaansoek as afgekeur of goedgekeur).
- Voorspelling (bv. voorspel die prys van 'n sekere aandeel op 'n sekere tyd).
- Beraming (bv. beraam die verwagte respons op 'n direkte bemarkingspoging van 'n sekere produk).
- Beskrywing of modellering (bv. beskryf die verwantskap tussen veranderlikes).

Kennisblootleggingstegnieke is in wese wiskundige metodes wat op die rekenaar geïmplementeer word om probleme op te los.

### 4.3. Uitdagings vir die bedryfswiskundige

As gevolg van die ontwikkeling in die rekenaartegnologie en kommunikasie-middele word massas data deur ondernemings versamel. In Suid-Afrika is daar baie voorbeelde, byvoorbeeld *ABSA*, *Telkom* en *SANLAM* om 'n paar te noem. Volgens Hand (1998) word massiewe databasisse in die VSA by ondernemings soos byvoorbeeld *Wal Mart*, *AT&T*, *Mobil Oil* en *NASA* gevind. Hy noem byvoorbeeld dat *Wal Mart* 20 miljoen transaksies daaglik hanteer, *AT&T* hanteer 200 miljoen oproepe per dag, *Mobil Oil* berg 100 terragrepe data oor oliesoektogte, en *NASA* skat dat hul *Earth Observing System* 50 gigagrepe data per uur aan die begin van die 21ste eeu sal insamel. Huber(1997) gee 'n breedweg 'n klassifikasie van datagrootte as *baie klein* ( $10^2$  grepe), *klein* ( $10^4$  grepe), *gemiddeld* ( $10^6$  grepe), *groot* ( $10^8$  grepe), *baie groot* ( $10^{10}$  grepe) en *massief* ( $10^{12}$  grepe).

Dit alles het tot gevolg dat 'n groot behoefte ontstaan ten opsigte van die outomatiese analise van data om patrone of strukture te vind wat van besigheids- of wetenskaplike belang kan wees.

Fayyad, Djorgovski & Weir (1996) kyk vanuit die oogpunt van die Rekenaarwetenskap en stel hierdie uitdaging soos volg: *We face a critical need for information processing technology and methodology with which to manage this data avalanche in order to produce interesting scientific results quickly and*

*efficiently. Developments in the fields of knowledge discovery in data bases, machine learning and related areas can provide at least some solutions.*

Hand (1998) kyk vanuit die oogpunt van Statistiek en stel die uitdaging so: *To many statisticians the notion of automatic data analysis (with little control by the researcher) will be abhorrent. Data analysis is as much an art as a science. However, the imperatives of the sheer volume of data mean that we have no choice.*

Huber (1997), ook 'n statistikus, stel die uitdaging só: *The human ability to inspect a data set, or even a meaningful part of it, breaks down far below terabyte sizes. I believe that attempts to circumvent this by automating some aspects of exploratory data analysis are futile. The available success stories suggest that the real function of data mining and knowledge discovery in data bases is not machine discovery of interesting structures by itself, but targeted extraction and reduction of data to a size and format suitable for human inspection.*

Kennisblootlegging ten opsigte van groot databasisse is duidelik 'n vakgebied in wording soos deur die volgende aanhaling van Fayyad (1998) blyk. *It is extremely important to remember that the achievements so far, while very useful and much needed, are still far from the general expectations being set in the public's mind. We are yet to have general data mining solutions and an understanding of the fundamental problems in our field and how they can be approached systematically. We also lack enough techniques to adequately cover the set of problems we face.*

Uit bostaande is dit duidelik dat die analise van hierdie massas data baie uitdagings vir Bedryfswiskunde en Informatika en meer spesifiek vir statistici en rekenaarwetenskaplikes bied. Meer spesifieke uitdagings sal onder die hofies data, data-hantering, algoritmes en statistiese beginsels bespreek word.

### **Uitdagings betreffende data-analise**

Statistiese data-analise behels die intensiewe analise van 'n enkele datastel. Wanneer met massiewe datastelle gewerk word, word intensiewe persoonlike analises onmoontlik om uit te voer a.g.v. die geweldige volume data. Vooraf-analise van data met behulp van spreidingsdiagramme mag selfs van geen nut wees nie, want hoe interpreteer iemand 'n miljard datapunte op een grafiek? Daar is min ander uitweë as om swaar te leun op rekenaarsteun wat self patrone in data soek met relatief min ingryping of tussenkoms van die analis.

Groot stelle data het, behalwe baie rekords of datapunte, gewoonlik baie veranderlikes. Dit lei tot die sogenaamde vloek van dimensionaliteit (Engels: *curse*

of *dimensionality*), byvoorbeeld met 1000 binêre veranderlikes is daar in die orde van  $10^{300}$  selle, 'n getal wat selfs 'n miljard rekords na min data laat lyk.

Statistiek het meesal te doen met die analise van numeriese data. Databasisse bevat al hoe meer andersoortige data, soos beelddata, klankdata, teksdata en geografiese data. In dié verband is 'n huidige interessante onderwerp die ontginning van kennis op die Internet. Groot uitdagings bestaan om prosedures te skep wat dit moontlik sal maak om bogenoemde tipes data sinvol te kan hanteer.

Groot datastelle is tipies heterogeen met baie interne struktuur, wat baie verskillend is van die datastelle wat gewoonlik deur statistici ontteed word.

### **Uitdagings betreffende datahantering**

'n Groot datastel kan oor verskillende kleiner databasisse verspreid wees. 'n Goeie voorbeeld is wanneer opnamedata met reeds versamelde data verbind moet word ten einde gevolgtrekkings te maak. Die verspreide databasisse lei gewoonlik tot inkonsekwentheid en bemoelijk standaard statistiese prosedures soos steekproefftrekking.

Skoon data is 'n nodige voorvereiste vir die meeste statistiese analises. In groot datastelle is dit redelik seker dat data op die een of ander manier ongeldig sal wees, veral as data versamel is waar moontlike menslike wisselwerking ter sprake is.

Steekproeffoute in groot databasisse kan dramatiese effekte hê. As 1% van die data uit 'n ander verdeling getrek word as die onderliggende verdeling van die populasie en die steekproef bestaan uit 'n miljard rekords, impliseer dit 'n miljoen rekords uit die ander verdeling.

### **Uitdagings betreffende algoritmes**

Gedurende modelpassing mag dit gebeur dat alle data nie in die hoofgeheue van die rekenaar sal pas nie. Dit sal beteken dat aanpassingsprosedures of sekwensiële tegnieke ontwikkel moet word om modelle te pas.

Die meeste standaard statistiese tegnieke is gebaseer op die aanname dat data-items uit dieselfde verdeling afkomstig is. Verbreking van die onafhanklik identies verdeelde aanname is waarskynlik die norm wanneer met groot datastelle gewerk word. Dit is baie waarskynlik dat sekere dele van die steekproefruimte meer verteenwoordig as ander is.

Opwaartse bywerking van sekere algoritmes word moeilik. Byvoorbeeld, konvensionele trosanalise-algoritmes kan nie massiewe datastelle hanteer nie.

Data word tipies voortdurend ingesamel, wat beteken dat databasisse gedurig besig is om te verander. Die behoefte om hierdie data in reële tyd te prosesseer en te analiseer word al hoe groter. Dit verskil baie van die gewone situasie wanneer statistiese analises uitgevoer word. Die feit dat vinnige antwoorde verlang word en die grootte van datastelle, lei tot moeilike vrae oor die doeltreffendheid van statistiese algoritmes.

Ten einde outomatiese data-analise verder te voer is meer algoritmes nodig soos byvoorbeeld modelseleksie-algoritmes in regressie en klassifikasiebome.

### **Uitdagings vir Statistiek**

Statistici het tipies in omstandighede gewerk waar daar 'n tekort aan data is in stede van 'n oorvloed. Dit het gelei tot ontwikkeling van prosedures waar die Tipe1-fout op 'n redelike drempel vasgemaak is en genoegsame data dan versamel is om voldoende onderskeidingsvermoë te gee betreffende toepaslike alternatiewe hipoteses. Wanneer 'n oorvloed data beskikbaar is, is hierdie strategie betwisbaar. Die resultate van sulke toetse kan lei tot sterk aanduidings van selfs baie klein effekte wat van geen praktiese waarde is nie. In die plek van statistiese betekenisvolheid behoort gekyk te word na substantiewe betekenisvolheid, m.a.w. is die effek belangrik of waardevol of nie?

Huber (1997) maak die volgende interessante waarneming: *Huge data sets usually are not just more of the same, they have to be huge because they are heterogeneous, with more internal structure, such that smaller sets would not do. As a consequence, sub-samples and techniques based on them, like the bootstrap, may no longer make sense.*

Aannames van baie statistiese prosedures is op stasionariteit gebaseer. Populasieskuiwing (Engels: *Population drift*) kan gebeur omdat die onderliggende populasie verander (byvoorbeeld die populasie van leningaansoekers kan verander soos die ekonomie verhit en afkoel). Populasieskuiwing kan onopgespoorbly as die vasleggingstyd van individuele rekords nie opgeneem is nie.

Seleksiesydigheid of die verwringing van die geselekteerde steekproef weg van die gewone eenvoudige ewekansige steekproef is 'n belangrike onderskatte probleem wanneer met groot stelle data gewerk word. Tipies sal sulke datastelle gemaklikheid- of geleentheidsteekproewe wees eerder as die geïdealiseerde

ewekansige steekproewe. Die mate waarin dit die analise sal beïnvloed, hang af van die doelwitte van die analise, veral as inferensie aangaande die onderliggende populasie gemaak wil word. 'n Voorbeeld is die ontwikkeling van telreëls (Engels: *Scoring rules*) om leningaansoekers te evalueer. Dikwels is omvattende data net beskikbaar betreffende dié wat aansoek gedoen het en as "goed" geklassifiseer is. Dié wat as "sleg" gegradeer was, sou afgekeur gewees het met geen kennis van hul ware status nie. Indien sulke data gebruik word om afleidings of gevolgtrekkings te maak oor toekomstige gedrag van aansoekers, kan growwe foute begaan word. Wat nodig is, is uitgebreide modelle wat die onderliggende steekproefseleksiemechanisme in ag neem.

Pogings om data-analise te outomatiseer ten einde groot datastelle te hanteer, sal met hoë waarskynlikheid lei tot die identifikasie van vals of toevallige datakonfigurasies as patrone of strukture. Dit is verwant aan die oormatige passingsprobleem (Engels: *overfitting problem*) en prosedures moet gevind word wat hierteen kan waak. Hand (1998) bied 'n paar gedagtes in dié verband, nl.

- die gebruik van telreëls in plaas van waarskynlikhede,
- beperking van die familie van modelle wat vir passing oorweeg word,
- die optimisering van 'n gepenaliseerde pasgehaltefunksie en
- die krimpings van 'n oorgepaste model.

Dit is belangrik om te onthou dat die vakkundige sal besluit of 'n geïdentifiseerde patroon van belang is en nie die statistiese struktuur nie.

Die analise van 'n enkele homogene datastel deur middel van 'n statistiese tegniek kan beskou word as taktiese data-analise. In groot datastelle word strategie al hoe belangriker, dit wil sê om te besluit watter subgroep deur middel van watter tegniek geanaliseer behoort te word. Modelseleksieprocedures word dus van groot belang en ander prosedures wat data-analisemetodologie ondersteun.

## 5. Slot

In hierdie rede het ek Bedryfswiskunde en Informatika, beroepsmoontlikhede daarin en uitdagings bespreek. Myns insiens is Bedryfswiskunde en Informatika nie 'n vak in eie reg nie maar 'n wyse van opleiding in die tradisionele ekonomiese, wiskundige en verwante rekenaarwetenskaplike vakgebiede. Die opleiding poog om die vakrigtings op so 'n wyse te integreer dat die effek van die integrasie groter is as die som van die effekte van die dele van vakgebiede waarop dit gebaseer is.

Die waardetoevoeging geskied veral met die oog op 'n professionele loopbaan in die finansiële industrie en in die besonder die loopbaan van finansiële ingenieur. Die tipe bedryfsgerigte opleiding moet egter op so 'n wyse geskied dat dit die markwaarde van tradisionele beroepe soos dié van operasionele navorser en statistikus in die proses verhoog. Omdat Bedryfswiskunde en Informatika bedryfsgerig is, sal die tipe probleme wat hoofsaaklik in die finansiële industrie ervaar word, teruggevoer word na die tradisionele vakke wat nuwe uitdagings sal skep. So 'n voorbeeld is die behoefte aan kennisblootlegging van groot databasisse. Dit is op die oomblik 'n daadwerklike behoefte wat fundamentele verskille het met die basiese onderliggende teoretiese beginsels waarop die vakgebied Statistiek geskoei is, byvoorbeeld homogene data en kanssteekproewe is die uitsondering eerder as die reël en die toepassing van waarskynlikheidsleer is in baie gevalle onvanpas. 'n Paradigmaverskuiwing in die tradisionele statistiek en gesamentlike navorsing met kunsmatige intelligensie is nodig om die teoretiese en filosofiese onderbou van die nuwe vak te verskaf.

### Verwysings

- Ackoff, R.L. (1979). *The Future of Operations Research is Past*. Journal of the Operational Research Society, Vol. 30, No. 2, pp. 93-104.
- Botha, S.; Gryffenberg I.; Hofmeyr, F.R.; Lausberg, J.L.; Nicolay, R.P.; Smit, W.J.; Uys, S.; Van der Merwe, W.L. and Wessels, G.J. (1997). *Guns or Butter: Decision Support for Determining the Size and Shape of the South African National Defense Force*. Interfaces, Vol. 27, No. 1, pp. 7-28.
- Dalton, G.W. and Thompson, P.H. (1986) *Strategies for Career Management*. New York: Scott, Foresman.
- de Jongh, P.J. and de Wet, T. (1993). *Statistics the High Road and the Low Road*. Ongepubliseerde manuskrip.
- de Jongh, P.J.; Carden, K.J. and Rogers, N.A. (1994). *Future: A Knowledge Based System for Threat Assessment*. Interfaces, Vol. 24, No. 2, pp. 76-86.
- Erasmus, C.M. (1999). Inouguerele rede. In proses van voorbereiding.
- Fayyad, U.M.; Djorgovski, S.G. and Weir, N. (1996). *Automating the Analysis and Cataloging of Sky Surveys* in Advances in Knowledge Discovery and Data Mining, eds Fayyad and Piatetsky-Shapiro.



- Fayyad, U.M.(1998). *Editorial column*. Data Mining and Knowledge Discovery, Vol. 2, No. 2, pp.115-119.
- Gass, S.I. (1996). *Model World: On Academics, Applications, and Publications*. Interfaces, Vol. 26, No. 6, pp. 105-111.
- Hahn, G. and Hoerl, R (1998). *Key Challenges for Statistics in Business and Industry (with discussion)*. Technometrics, Vol. 40, No. 3, pp. 195-213.
- Hand, D.J. (1998). *Data Mining: Statistics and More?* The American Statistician, Vol. 52, No. 2, pp. 112-118.
- Huber, P.J. (1997a). *Strategy Issues in Data Analysis*. Proceedings of the Conference on Statistical Science Honoring the Bicentennial of Stefano Franscini's Birth, eds C. Malaguvia, S. Morgenthaler, and E. Ronchetti, Basel: Birkhauser-Verlag.
- Huber, P.J. (1997b). *From Large to Huge: A Statistician's Reactions to KDD and DM*. Ongepubliceerde manuskrip.
- Kettenring, J.R. (1997). *Shaping Statistics for Success in the 21st Century*. Journal of the American Statistical Association, Vol. 92, No. 440, pp. 1229-1234.
- Kersten, H.M.P. (1993). *Het kwantificeren van bedrijfsprocessen: essentieel voor flexibel, veilig en optimaal ondernemen*. Rede uitgesproken bij de aanvaarding van het ampt van bijzonder hoogleraar bedrijfswiskunde bij de Faculteit der Wiskunde en Informatica van de Vrije Universiteit te Amsterdam op 12 Februari 1993.
- Woolsey, R.D. and Maurer, R.A. (1995). *The fifth column: On the proper training of future management II*. Interfaces, Vol. 25, No. 2, pp. 74-80.