



Markku Suni

**SASiin ja sassiin
meillekin SAS**



TURUN AMMATTIKORKEAKOULU
TURKU UNIVERSITY OF APPLIED SCIENCES

É L A L
PARIS

Jopa **12x ENEMMÄN**
NÄYTTÄVYYTTÄ, ilman paakkuja

UUSI MASKARA

EXTRA-VOLUME
COLLAGENE

HYDRA-COLLAGENE
RAKENTAA VOLYYMIA

EXTRA VOLUME HARJA
2 X SUUREMPI KUIN
PERINTEINEN HARJA

Koska olet sen arvoinen
L'ORÉAL
PARIS

BEA KOULU
D SCIENCES

Miten tuo lasketaan ?

Tietysti SAS-välinein

Siksi tarvitaan SAS

Ja taitoa sen käyttöön

Ja niin SAS saatiin Turun ammattikorkeakouluun

Mihin sitä voitaisiin käyttää?

Miten saataisiin väki tehokkaimmin oppimaan ?

Onhan niitä kursseja eri muotoisia.

Onhan meillä opettajia omastakin takaa

Ja peräti yksi SASia ennestään tunteva

Nopeita tuloksia:

pikakurssi

Tehokkaasti tuloksia:

intensiivikurssi

Näiden summana:

pikainen intensiivikurssi

Kesäkuun alussa kaksi viikkoa SAS-peruskurssia

Joka arkipäivä 9 - 16

Opettaja paikalla 9 - 13

Harjoitustehtävillä jatketaan

Opetuskielenä englanti

Pikaisen intensiivikurssin sisältö

Perusteita, periaatteita

Raportointi: SORT, PRINT, WHERE, OPTIONS

Raportointi: FORMAT, PROC FORMAT

Raportointi: TABULATE,

Raportointi: muut proseduurit

Ohjelmointi: DATAvaihe, INPUT ja sen variaatiot

Ohjelmointi: SAS-funktiot, merkkijonot, aikakäsite

Ohjelmointi: DATAvaiheen muut lauseet

Ohjelmointi: Tiedostot ja niihin liittyvät optiot

Tässä meni ensimmäinen viikko

Pikaisen intensiivikurssin sisältö – toinen viikko

Ohjelmointi: Tiedostot, MERGE variaatioineen

Ohjelmointi: Tiedostot, automaattimuuttujat

Ohjelmointi: Velhojen tiedostometkut

Erillinen esitys: vierailija yrityksestä

Ohjelmointi: Tietokantayhteydet ja SQL

Ohjelmointi: Makrokielen alkeet lapsille

Ohjelmointi: Makrokieli aikuisille

Ohjelmointi: Dynaaminen ohjelmointi, velhojen makrometkut

Pikainen intensiivikurssi oli todella juuri sitä

- Tämä kurssi oli todella pikainen
- Se oli raskas opettajalle ja opiskelijalle
- Materiaalia oli paljon ja tekemistä myös
- Mutta useimmat oppivat hyvin ja paljon
- Tällainen koulutusmuoto on hyvä toisinaan
 - motivoitunut opettaja
 - motivoituneet opiskelijat
 - valmius tehdä työtä ja nähdä vaivaa
 - kunnollinen ympäristö: laitteet, materiaali

Niin, mutta tämähän on koulu . . .

- Oppi on myös osoitettava jotenkin
- Perinteinen tapa on lopputentti
- Sensijaan tehtiinkin toisin
- Jokainen osallistuja velvoitettiin kokoamaan henkilökohtainen SAS-käsikirja
 - Tokihan osaaja osaa kertoa muillekin

Niin, mutta tämähän on koulu . . .

- Jokainen osallistuja kokoaa SAS-käsikirjan
 - Kyseessä ei ole ”perinteinen” portfolio
 - Asetelma: kyseinen osallistuja on firmansa ainoa SAS-osaaja ja joutuu kouluttamaan muita
 - Tämä käsikirja tulee muiden materiaaliksi
 - Pitkään konsulttina ja koodarina toiminut opettaja arvioi sen sisältöä tämän näkökohdan mukaan

Niin, mutta tämän on koulu . . .

- Ja palautetta antoivat osallistujatkin:

Mitä tulee itse SAS-kurssiin, on siitä minulla vain yksi sana. Mahtava! Se oli paras kurssi mitä näin ensimmäisenä vuotena on tullut käytyä. Paljon asiaa, mentiin eteenpäin, oli mielenkiintoista. Vaadittiin myös paljon (toki jos vaan jaksoi tehdä, mitä ei kaikki näyttäneet niinkään sisäistävän). Mutta kurssina oli todella loistava.

Yritysvierailija oli hyvä - tästä voisi ottaa malliksi muillekin kursseille.

Niin, mutta tämän on koulu . . .

- Ja palautetta antoivat osallistujatkin:

I had never heard of SAS Programming before this course. This course has taught me a lot about SAS software including its products, applications and uses. It was a very educating and interesting course and has been an eye opener to me by revealing an integrated software that can perform so many tasks in all areas of life and businesses. *I never thought such a program could exist.* It would have been very unfortunate for me graduating as an IT Engineer without having a basic knowledge about SAS. I would suggest that if possible, these SAS programming course be included in our study program because of its usefulness.

Voihan SASia käyttää oikein työhönkin . . .

- Sain tehtäväkseni analysoida keskeyttämisiä

Miksi opiskelijat keskeyttävät opiskelunsa
Tietoliikenteen ja sähköisen kaupan koulutusalueella
Turun ammattikorkeakoulussa?

Menetelmänä:

Business Intelligence

Tiedon jalostus

Voihan SASia käyttää oikein työhönkin . . .

Miten tutkitaan ?

- menetelmä vastaa liikeyritysten tutkimustapaa:
 - miksi asiakkaat luopuvat meidän palveluistamme?
- lähteenä yksinomaan opintorekisteri
 - ei yritystäkään haastatella opiskelijoita
- saadaan vain se tieto, jota rekisterissä on
 - siihen voidaan palata uudelleen ja uudelleen

Voihan SASia käyttää oikein työhönkin . . .

Kun kysytään ihmisiltä jotain
Mitä saadaan vastaukseksi?

- Mitä kysytään?
- Mitä kysymys tarkoittaa; ymmärtävätkö molemmat sen samoin?
- Mitä kysyjä todella haluaa saada selville?
- Mitä vastaaja olettaa kysyjän haluavan saada selville?
- Mitä on mahdollista saada selville?

Nämä asiat voivat poiketa toisistaan paljonkin.

Lähtökysymys: keskeyttäneiden määrä

BI-tutkija kehittää joukon jatkokysymyksiä:

- kuinka paljon suoritettu ennen keskeyttämistä?
- mitä kursseja suoritettu ennen keskeyttämistä?
- onko tässä suhteessa jotain yhtenäisyyttä?
- mikä kurssi olisi ollut seuraavaksi edessä?
- onko olemassa jotain "pelättyjä" kursseja
- onko jotain muuta tapaa ryhmitellä keskeyttäneitä?
- onko jokin "vaarallinen" opiskelijatyyppe?
- onko jokin tietty kurssiyhdistelmä?
- onko jokin oppimisen määrä, joka tuntuu riittävän?

Menetelmät, välineet, aineistot

Materiaali: kaksi Turun AMK:n opintorekisteriotosta

- Vuoden 2003 aikana eronneet TSK:n opiskelijat
- Vuoden 2004 aikana eronneet TSK:n opiskelijat

Laskettiin lisäksi:

- opiskelijan suorittamat opintoviikot
- opiskeluaika
- aika ennen ensimmäistä suoritusta
- aika viimeisen suorituksen jälkeen
”virkaikä”
- aika sillä välillä
- keskimääräinen suoritustahti

Ongelma

Materiaali ei ollut kovin suuri.

→ Oli vaikea saada mitään selviä eroja

→ Voidaan siis todeta, että mitään ei voida todeta ???

Päätin siirtyä sumean logiikan käyttöön

Asia ei ole vain tosi tai väärä.

On totuuden aste (degree of truth)

Esim. Nuori nainen, vanha mies, ahkera opiskelija, . . .

Päätely tapahtuu laskemalla totuusarvoja

Sumea logiikka antaa joustavuutta

Jäljittelee ihmisen päättelyä

Ongelma

Materiaali ei ollut kovin suuri.

→ Oli vaikea saada mitään selviä eroja

→ Voidaan siis todeta, että mitään ei voida todeta ???

Päätin siirtyä sumean logiikan käyttöön

Asia ei ole vain tosi tai väärä.

On totuuden aste (degree of truth)

Esim. Nuori nainen, vanha mies, ahkera opiskelija, . . .

Päätely tapahtuu laskemalla totuusarvoja

Sumea logiikka antaa joustavuutta

Jäljittelee ihmisen päättelyä

Ongelma

Materiaali ei ollut kovin suuri.

→ Oli vaikea saada mitään selviä eroja

→ Voidaan siis todeta, että mitään ei voida todeta ???

Päätin siirtyä sumean logiikan käyttöön

Asia ei ole vain tosi tai väärä.

On totuuden aste (degree of truth)

Esim. Nuori nainen, vanha mies, ahkera opiskelija, . . .

Päätely tapahtuu laskemalla totuusarvoja

Sumea logiikka antaa joustavuutta

Jäljittelee ihmisen päättelyä

Sumea logiikka – sumea luokitus

Määrittelin materiaaliin sumeita muuttujia:

- ALKUPITKA** pitkä aika ennen ens. suorituksen kirjaamista
- PALJON_OV** suorittanut paljon opintoviikkoja ylipäänsä
- VAHAN_OV** suorittanut vähän opintoviikkoja ylipäänsä
- LOPPUPITKA** pitkä aika viimeisen suorituksen ja lopetuksen välillä
- PIAN** eronnut nopeasti, lyhyen ajan kuluttua
- KAUAN** opiskellut kauan ennen keskeytystä
- KESKHYVA** keskimäärin suorittanut paljon opintoviikkoja (vuodessa)



Sumea logiikka - sumea luokitus

Määrittelin materiaaliin sumeita muuttujia:

Sumeuttaminen tapahtui suhteellisesti ja lineaarisesti:

maksimiarvo → totuusarvo = 1

minimiarvo → totuusarvo = 0

välillä olevat arvot → totuusarvo lineaarisesti
interpoloiden

Huom: ei ole voimassa: $PALJON_OV = 1 - VAHAN_OV$



Sumea logiikka - sumeat luokitusmuuttujat

Määrittelin materiaaliin sumeita luokitusmuuttujia:

```
luovuttaja = ( pian    + vahan_ov    ) / 2.0;
             /* lopetti pian, suoritti vähän */
ahkera      = ( kauan   + paljon_ov  ) / 2.0;
             /* jatkoi kauan, suoritti paljon */
laiska      = ( kauan   + vahan_ov    ) / 2.0;
             /* jatkoi kauan, suoritti vähän */
sinnikas    = ( kauan   + keskhyva   ) / 2.0;
             /* jatkoi kauan, keskim. paljon */
hiipuja     = ( kauan   + loppupitka ) / 2.0;
             /* jatkoi kauan, pitkä loppuaika */
nopea       = ( pian    + paljon_ov  ) / 2.0;
             /* nopeasti opiskellut, paljon */
```



Sumea logiikka - sumeat luokitusmuuttujat

Määrittelin materiaaliin sumeita luokitusmuuttujia:

```
hyvahiipui = ( kauan+ loppupitka+ paljon_ov) / 3.0;
    /* jatkoi kauan, suoritti paljon, pitkä loppuaika */
alkuhyvin  = ( ( 1 - alkupitka ) + paljon_ov + kauan ) / 3.0;
    /* aloitti riuskasti, suoritti paljon, jatkoi kauan */
saivauhtii = ( alkupitka + paljon_ov + kauan ) / 3.0;
    /* aloitti hitaasti, suoritti paljon, jatkoi kauan */
```



Sumea logiikka – sumeat luokitusmuuttujat

Haluttaessa luokitella opiskelijat näiden luokitusmuuttujien mukaisiin luokkiin voidaan kiinnittää luokkien rajat: 0, 0.2, 0.4, 0.6, 0.8, 1
→ Luokkarajat voivat osoittautua jotenkin huonoiksi

Materiaalia en tunne ennestään.

Halusin määritellä luokat ja niiden rajat sumeiksi joustavasti.



Sumea logiikka – sumeat luokitusmuuttujat

Miten määritellä luokkarajat sumeasti tietojen perusteella

- **Sumea hermoverkko** olisi hyvä väline
- Kohosen **SOM** (Self Organizing Maps)
- **LLSQ-menetelmä** arvioimaan luokkiin sopimista

Näiden yhdistelmänä ja modifikaationa kehitin oman algoritmin.

Pomohan on oikeassa.



Sumea logiikka – sumeat luokitusmuuttujat

1. asetetaan neljä tasavälistä pistettä (arvojoukko kuuteen luokkaan)
2. Lasketaan joka havaintoarvon etäisyys jokaiseen pisteeseen
3. Pienin etäisyys jätetään jäljelle ja vastaavan pisteen numero. Myös jätetään +1 tai -1 (etäisyyden etumerkki)
4. Kun kaikki havainnot on käsitelty, lasketaan etäisyyksien neliösumma ja kutakin pistettä kohti + ja - ykkösten summa
5. Siirretään kutakin pistettä 0.01 ykkösten summan suuntaan
6. Toistetaan vaiheet 2 - 4
7. Jos neliösumma pieneni, otetaan äskeinen siirros todella käyttöön
Jos neliösumma ei pienentynyt, lopetetaan
8. Joka tapauksessa lopetetaan N:n askeleen jälkeen
9. Saadut luokitukset tilastoidaan



Sumea logiikka – sumeat luokitusmuuttujat

Päädyin luokkamuuttujien arvoihin

1 = 'Ei yhtään'

2 = 'Hyvin vähän'

3 = 'Vähän'

4 = 'Enemmän'

5 = 'Paljon'

6 = 'Varma';

1 = 'Varma'

2 = 'Paljon'

3 = 'Enemmän'

4 = 'Vähän'

5 = 'Hyvin vähän'

6 = 'Ei yhtään';

Kumpaa käytin, riippui ko. arvon suunnasta



Sumea luokitus tarkastelun kohteena

| Saivauhtii | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|-------------|-----------|---------|----------------------|--------------------|
| Hyvin vähän | 95 | 59.01 | 95 | 59.01 |
| Vähän | 33 | 20.50 | 128 | 79.50 |
| Enemmän | 33 | 20.50 | 161 | 100.00 |

Vain 20% sai vauhtia.

Antaako alku viitteitä jatkosta?

Tässä näyttää siltä



| laiska | Koulohj | | Total |
|-------------|---------|-----------|--------|
| Frequency | | | |
| Percent | | | |
| Row Pct | | | |
| Col Pct | SALOTSK | TEKNIikka | |
| Paljon | 12 | 3 | 15 |
| | 7.45 | 1.86 | 9.32 |
| | 80.00 | 20.00 | |
| | 21.05 | 2.88 | |
| Enemmän | 8 | 11 | 19 |
| | 4.97 | 6.83 | 11.80 |
| | 42.11 | 57.89 | |
| | 14.04 | 10.58 | |
| Vähän | 22 | 67 | 89 |
| | 13.66 | 41.61 | 55.28 |
| | 24.72 | 75.28 | |
| | 38.60 | 64.42 | |
| Hyvin vähän | 15 | 23 | 38 |
| | 9.32 | 14.29 | 23.60 |
| | 39.47 | 60.53 | |
| | 26.32 | 22.12 | |
| Total | 57 | 104 | 161 |
| | 35.40 | 64.60 | 100.00 |



| laiska | Koulohj | | Total |
|-------------|---------|-----------|--------|
| Frequency | | | |
| Percent | | | |
| Row Pct | | | |
| Col Pct | SALOTSK | TEKNIikka | |
| Paljon | 12 | 3 | 15 |
| Enemmän | 8 | 11 | 19 |
| Vähän | 22 | 67 | 89 |
| Hyvin vähän | 15 | 23 | 38 |
| Total | 57 | 104 | 161 |
| | 35.40 | 64.60 | 100.00 |

| Statistic | DF | Value | Prob |
|-----------------------------|----|---------|--------|
| Chi-Square | 3 | 18.1357 | 0.0004 |
| Likelihood Ratio Chi-Square | 3 | 17.8702 | 0.0005 |
| Mantel-Haenszel Chi-Square | 1 | 6.4204 | 0.0113 |

Alle 1%ⁿ riskillä: eroa on Turun ja Salon välillä
 ➔ Laiskuus asuu Salossa

Yhteenveto

Sumea logiikka sopii hyvin tämänkaltaiseen työhön

- käyttö ei ole erityisen vaikeaa
- vaatii enemmän paneutumista oletuksiin ja sääntöihin
- antaa uusia luokitusmahdollisuuksia
- saadaan lukujen alta uutta "joustavaa" tietoa

SAS-ohjelmalla on helppo toteuttaa sumeaa päättelyä

SAS-välineet toimivat tässä erittäin hyvin



Kiitos osallistumisesta



TURUN AMMATTIKORKEAKOULU
TURKU UNIVERSITY OF APPLIED SCIENCES

www.turkuamk.fi