

Bayesian model validation metrics in retail datasets

Leevi Rönty

School of Science

Bachelor's thesis
Espoo 14.10.2020

Supervisor

Prof. Fabricio Oliveira

Advisor

DSc (Tech.) Mikko Ervasti



Aalto University
School of Science

Copyright © 2020 Leevi Rönty

The document can be stored and made available to the public on the open internet pages of Aalto University.
All other rights are reserved.



Author Leevi Rönty

Title Bayesian model validation metrics in retail datasets

Degree programme Engineering Physics and Mathematics

Major Mathematics and Systems Sciences

Code of major SCI3029

Teacher in charge Prof. Fabricio Oliveira

Advisor DSc (Tech.) Mikko Ervasti

Date 14.10.2020

Number of pages 14+2

Language English

Abstract

Your abstract in English. Keep the abstract short. The abstract explains your research topic, the methods you have used, and the results you obtained.

Keywords Bayesian models, model validation, information criterion, Facebook Prophet



Tekijä Leevi Rönty

Työn nimi Bayeslaisten mallien validointi vähittäismyynnin aineistoissa

Koulutusohjelma Teknillinen fysiikka ja matematiikka

Pääaine Matematiikka ja systeemitieteet

Pääaineen koodi SCI3029

Vastuopettaja Prof. Fabricio Oliveira

Työn ohjaaja TkT Mikko Ervasti

Päivämäärä 14.10.2020

Sivumäärä 14+2

Kieli Englanti

Tiivistelmä

Tiivistelmässä on lyhyt selvitys kirjoituksen tärkeimmästä sisällöstä: mitä ja miten on tutkittu, sekä mitä tuloksia on saatu.

Tämän opinnäytteen tiivistelmäteksti kirjoitetaan opinnäytteen luettavan osan lomakkeen lisäksi myös pdf-tiedoston metadataan \thesisabstract-makron avulla (kastoyllä). Kirjoita tähän luettavaan tiivistelmälomakkeeseen menevä teksti. Tässä saa olla erikoismerkkejä kuten kreikkalaiset kirjaimet ja rivinvaiho- ja kappaleenjako-merkit. Tämän tekstin on muuten oltava sama kuin metadatatiiivistelmän teksti.

Jos tiivistelmäsi ei sisällä erikoismerkkejä eikä kaipaa kappaleenjakoja, voit hyödyntää makroa \abstracttext luodessasi lomakkeen tiivistelmää (katso kommentti alla).

Avainsanat Bayeslaiset mallit, mallin validointi, informaatiokriteeri, Facebook Prophet

Contents

Abstract	3
Abstract (in Finnish)	4
Contents	5
1 Introduction	6
2 Aikaisempi tutkimus	6
2.1 Rakenne	6
2.2 Sivut ja kirjaintyypit	7
2.3 Asemointi	7
2.4 Numerointi	8
2.5 Lähdeviittausten käyttö	8
2.6 Lähdeluettelo	9
2.7 Muuta huomioitavaa lähdeluettelossa	12
3 Tutkimusaineisto ja -menetelmät	13
4 Tulokset	13
5 Yhteenveto	13
A Esimerkki liitteestä	15
B Toinen esimerkki liitteestä	16

1 Introduction

- Yritykset laativat paljon ennusteita - Mallien validointi on oleellinen osa prosessia - Bayeslaisten mallien validointi on kuitenkin suhteellisen uusi asia
 - Bayeslaisten mallien validoinnin nykytila on "unsatisfactory" - Kuhunkin käytetyimmistä menetelmistä liittyy ongelmia - Entä jos sakkofunktiona MAPE?
 - Mallinnetaan myyntiä Prophetilla - Implementoidaan IC:t prophetille - Vertailaan tuloksia 'ad-hoc' -menetelmiin
 - Ei välttisi pakollinen

2 Aikaisempi tutkimus

Tässä osassa selvitetään, mitä tutkimuksen kohteena olevasta aiheesta tiedetään entuudestaan. Selvityksen tulee kattaa tasapainoisesti koko tutkimuskenttä.

Kun opinnäytetyötä kirjoitetaan, on noudatettava ohjeita, jotka koskevat opinnäytteen rakennetta, käytäntöjä, muotoseikkoja sekä ulkoasua. Esitellään näitä ohjeita tarkemmin.

2.1 Rakenne

Opinnäytteen rakenteen tulee olla hyvän tieteellisen kirjoittamisen käytännön mukainen ja sisältää vähintään seuraavat osat:

1. Nimiölehti
2. Tiivistelmä
3. Sisällysluettelo
4. Symboli- ja lyhenneluettelo (ei yleensä käytetä systemissä)
5. Johdanto
6. Aikaisempi tutkimus. Työn luonteen niin vaatiessa otsikko voi olla myös »Teoreettinen tausta» tai näiden otsikoiden yhdistelmä.
7. Tutkimusaineisto ja -menetelmät. Otsikko voi olla myös jotain muuta, jos se sopii paremmin työn luonteeseen.
8. Tulokset
9. Tarkastelu. Työn luonteen niin vaatiessa otsikko voi olla myös »Johtopäätökset» tai »Yhteenvedo» tai edellä mainittujen otsikoiden yhdistelmä.
10. Lähteet
11. Liitteet.

Tiivistelmän ja symboli- sekä lyhenneluetteloiden väliin voi sijoittaa halutessaan esipuheen.

Työn osat 5–9 muodostavat *tekstiosan*. Työn yksittäisiä osia voidaan jakaa alaotsikoilla alaosiin, joita ei ole yllä esitetty. Alaotsikoiden käyttäminen selventää parhaimmillaan tekstiä, ja pahimmillaan sirpaloittaa sitä. Sirpaloitumista voi estää huolehtimalla siitä, että samalla sivulla ei esiinny useampaa alaotsikkoa. Tekstin jäsentelyssä on yleensä ongelmia, jos osassa on vain yksi alaosa, tai kirjoittaja joutuu käyttämään useampaa kuin kahta tasoa (osa ja alaosat): alaosien alaosat ovat harvoin tarpeen.

2.2 Sivut ja kirjaintyypit

Opinnäytteen tulee olla kirjoitettu koneella tai tekstinkäsittelyohjelmalla yksipuolisesti A4-kokoiselle paperille. Kandidaatintyön tekstiosan sopiva pituus on noin 15–20 sivua ja diplomityön noin 60 sivua. Työtä ei ole syytä tarpeettomasti pidentää.

Opinnäytteen tekstiosan kirjaintyyppin tulee olla antiikva eli serif-tyyppinen ja lisäksi kursivoimaton, lihavoimaton sekä kooltaan 12 pistettä (kuten tässä esityksessä). Groteskeja eli **Sans serif**-tyyppisiä kirjaintyyppejä (kuten Helvetica tai Arial) ei saa käyttää varsinaisessa tekstissä, mutta otsikoissa näitä voidaan käyttää. Otsikoissa voidaan käyttää kooltaan edellä mainittua suurempaa kirjaintyyppiä sekä tyylikeinoja, kuten lihavointia tai kursivointia. Tekstissä samantasoisten otsikoiden on kuitenkin oltava tyyliältään ja kirjainlajeiltaan yhteneväisiä.

Table 1: Taulukoissa ja kuvissa kirjaintyyppin voi valita tarkoituksenmukaisesti, mutta kuva- ja taulukkoteksteissä tulee käyttää samaa kirjaintyyppiä kuin varsinaisessa tekstissä. Huomaa taulukon numeroinnin sijoittuminen taulukon yläpuolelle.

A	1	$e^{j\omega t}$
B	2	$\Re(c)$
C	3	$a \in \mathbb{A}$

Opinnäytteen vasen marginaali (sidonnan puoli) on 35 mm ja oikea 25 mm. Ylämarginaali on 25 mm. Leipätekstin korkeus on enimmillään 230 mm. Tämän opinäytepohjan marginaalien pitäisi olla paperille tulostettuna oikein, mutta tulostimesta ja paperista riippuen voi esiintyä yhden tai kahden millimetrin suuruisia eroja.

2.3 Asemointi

Tekstiosan tekstissä käytetään kappaleiden erottamiseen sisennystä, mutta ensimmäistä otsikon, väliotsikon tai muun katkon jälkeistä kappaletta ei sisennetä. Jos kuva tai muu katko tulee kappaleiden väliin, suositellaan katkon jälkeisen kappaleen sisentämistä.

Mikäli oikea reuna halutaan tasata, tulee käyttää tavutusta ja lisäksi tarkistaa, ettei tekstiin jää lukemista häiritseviä pitkiä sanavälejä. Jos käytät opinnäytteen tekemisessä L^AT_EX-järjestelmää, tämä asia hoituu automaattisesti.

Opinnäytteen riviväli on 1, mikä on myös tämän opinnäytepohjan käytäntö. Kappaleiden tulee yleensä olla ainakin kolmen rivin pituisia, mutta myös liian pitkiä kappaleita tulee välttää. Tässä opinnäytepohjassa ei tekstin luonteen vuoksi voida täysin toteuttaa kappaleen pituutta koskevia vaatimuksia.

Yksittäisiä, kappaleen päättäviä tai aloittavia rivejä sivun alussa tai lopussa on vältettävä koko työssä, myös luetteloissa ja liitteissä.

2.4 Numerointi

Työn sivut numeroidaan siten, että kansilehti on ensimmäinen sivu ja numerointi on jatkuva työn loppuun asti. Sivunumeroinnissa käytetään arabialaisia numeroita.

Sivunumerointi alkaa tiivistelmä sivujen jälkeen esipuheesta, jos sellainen on tai sisällysluettelosivusta.

Lähdeluettelo alkaa uudelta sivulta. Lähdeluettelon sivunumerointi jatkuu viimeisestä tekstisivusta.

Jokainen liite alkaa uudelta sivulta. Liitteiden sivunumerointi jatkuu viimeisestä lähdeluettelon sivusta.

Sivunumero sijoitetaan sivun yläreunaan.

Matemaattiset kaavat numeroidaan arabialaisin numeroin. Kaavanumerointi ei saa katketa osien välissä (eikä niin tapahdukaan, jos käytät tätä opinnäytepohjaa). Kaikkia kaavoja ei tarvitse numeroida, vaan kirjoittaja voi käyttää harkintaa numeroinnin tarpeellisuudessa. Liitteissä olevat kaavat numeroidaan siten, että liitteen ajatellaan muodostavan numeroinnin kannalta itsenäisen ja yhtenäisen kokonaisuuden. Kaavan numero sijoitetaan oikealle puolelle alla olevan esimerkin mukaisesti

$$D(xy) = (Dx)y + x(Dy), \quad x, y \in \mathbb{A}. \quad (1)$$

Kaikki kuvat ja taulukot numeroidaan erillisen juoksevan numeroinnin mukaisesti kuten taulukosta 1 ja kuvista 1 ja 2 käy ilmi. Kuva 2 esittää MATLABilla piirrettyä käyrää. Liitteissä olevat kuvat ja taulukot numeroidaan siten, että liitteen ajatellaan muodostavan numeroinnin kannalta itsenäisen ja yhtenäisen kokonaisuuden. Liitteissä A ja B on esimerkkejä kaavojen (kaavat A1–A2 tai kaavat B1–B2), kuvien (kuva B1) ja taulukoiden (taulukko B1) numeroimisesta. Liitteet numeroidaan suuraakkosin (esimerkiksi Liite A, Liite B tai pelkästään A, B).

2.5 Lähdeviittausten käyttö

Lähdeviittaukset tulee tehdä huolellisesti ja johdonmukaisesti numeroviitejärjestelmän mukaisesti. Numeroviitteet järjestetään lähdeluetteloön viittausjärjestykseen, mutta jos lähdeluettelo on hyvin laaja (useita sivuja), järjestetään viitteet pääsanan mukaiseen aakkosjärjestykseen. Alaviitejärjestelmää ¹ ei käytetä.

Viitteen sijoittelussa noudatetaan seuraavia sääntöjä: Jos viite kohdistuu vain yhteen virkkeeseen tai virkkeen osaan, viite (Kauranen et al., 2006) sijoitetaan virkkeen sisään ennen virkettä päättävää pistettä. Jos taas viite koskee tekstin

¹Myöskään alaviitteenä olevia kommentteja ei suositella käytettäväksi.

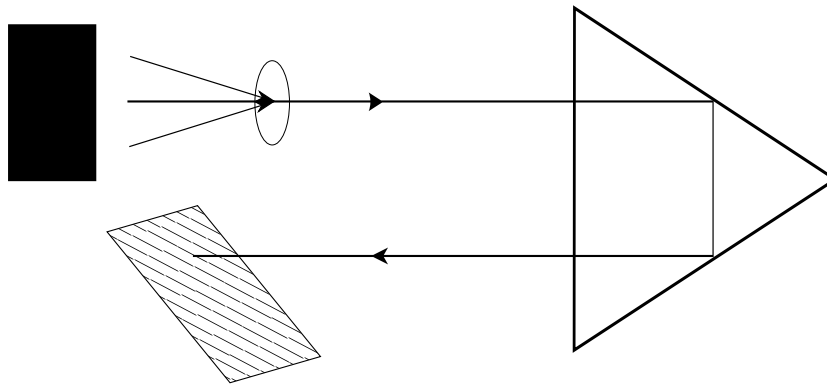


Figure 1: Tämä on esimerkki numeroidusta kuvatekstistä.

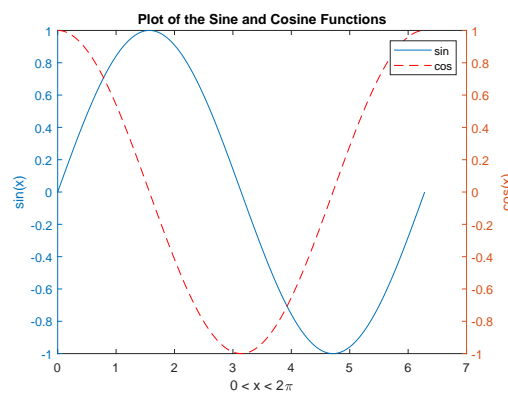


Figure 2: Tämä on esimerkki MATLAB-kuvasta.

useampaa virkettä tai kokonaista kappaletta, sijoitetaan viite kappaleen loppuun pisteen jälkeen. (Kauranen et al., 2006)

2.6 Lähdeluettelo

Lähdeluettelossa esiintyy tavallisesti seuraavassa esitettäviä lähteitä, joista on numeroviitejärjestelmässä ilmoitettava asianomaisessa kohdassa vaaditut tiedot.

Kirjasta ilmoitetaan seuraavat tiedot:

- tekijät
- julkaisun nimi
- painos, jos useita
- kustannuspaikka
- julkaisija tai kustantaja
- julkaisuaika
- mahdollinen sarjamerkintö.

Viitteet ([Kauranen et al., 2006](#); [Koblitz, 1994](#)) ovat esimerkkejä kirjan esittämisestä lähdeluettelossa. Viite ([Koblitz, 1994](#), s. 83–124) on esimerkki lähdeluettelossa esiintyvän kirjan tiettyjen sivujen esittämisestä tekstissä.

Artikkelista kausijulkaisussa ilmoitetaan seuraavat tiedot:

- tekijät
- artikkelin nimi
- kausijulkaisun nimi
- julkaisuvuosi
- kausijulkaisun volyymi tai ilmestymisvuosi
- kausijulkaisun numero
- sivut, joilla artikkeli on.

[Bardeen et al. \(1957\)](#) ja [Deschamps \(1981\)](#) ovat kirjoittaneet artikkelit, joita käytetään tässä esimerkkeinä artikkelin esittämisestä lähdeluettelossa.

Kokoomateoksen luvusta tai osasta ilmoitetaan seuraavat tiedot:

- luvun tai osan tekijät
- luvun tai osan nimi
- maininta »Teoksessa»
- koko teoksen toimittajat sekä maininta »(toim.)»
- koko teoksen tai konferenssin nimi
- konferenssiesitelmän kyseessä ollessa sen pitopaikka ja -aika
- painos, jos useita
- kustannuspaikka
- julkaisija tai kustantaja, jos aihetta tämän ilmoittamiseen on
- julkaisuaika
- sivut, joilla luku tai osa on
- mahdollinen sarjamerkintä.

Viitteet ([Sihvola et al., 1998](#); [Lindblom-Yläne and Wager, 2002](#)) ovat esimerkkejä kokoomateoksen luvun tai osan esittämisestä lähdeluettelossa.

Opinnäytetyöstä ilmoitetaan seuraavat tiedot:

- tekijä
- työn nimi
- opinnäytetyön tyyppi
- oppilaitoksen nimi
- oppilaitoksen sijaintipaikka
- vuosiluku.

Viitteet ([Miinusmaa, 1977](#); [Lönnqvist, 2006](#)) ovat esimerkkejä opinnäytteen esittämisestä lähdeluettelossa.

Haastattelusta ilmoitetaan seuraavat tiedot:

- haastatellun henkilön nimi
- haastatellun henkilön arvo tai asema
- haastatellun henkilön edustama organisaatio
- organisaation osoite
- maininta siitä, että kyseessä on haastattelu ja haastattelun päivämäärä.

Palmgrenin ([2007](#)) haastattelu on esimerkki haastattelun esittämisestä lähdeluettelossa.

Osa sähköisessä muodossa olevista artikkeleista on saatavissa myös painettuina. *Vain verkosta saatavissa olevasta artikkelista* esitetään seuraavat tiedot:

- tekijät
- artikkelin nimi
- kausijulkaisun nimi
- viestintyyppi
- laitos tai volyymi
- kausijulkaisun yksittäistä osaa koskeva merkintä tai numero
- julkaisuvuosi tai maininta »Päivitetty» ja päivitysaika
- maininta »Viitattu» ja viittaamisen ajankohta
- maininta »Saatavissa» ja URL tai maininta »DOI» ja DOI-numero (DOI=Digital Object Identifier).

Viitteet ([Ribeiro et al., 2006](#); [Stieber, 2006](#); [Pohjois-Koivisto, 2005](#)) ovat esimerkkejä sähköisessä muodossa olevan artikkelin esittämisestä opinnäytteen lähdeluettelossa. Viitteet ([Ribeiro et al., 2006](#)) ja ([Stieber, 2006](#)) ovat saatavissa sekä painettuna että verkosta, joten viitteiden esitystapa mukailee painetun artikkelin viitteen esitystapaa. Viite ([Pohjois-Koivisto, 2005](#)) on saatavissa vain verkosta ja siitä esitetään yllä vaaditut tiedot.

Valitettavasti sähköisessä muodossa olevasta artikkelista ei ole aina saatavissa laitos-, volyymi- tai numerotietoja.

Sähköisessä muodossa olevasta opinnäytetyöstä ilmoitetaan seuraavat tiedot:

- tekijä
- työn nimi
- opinnäytetyön tyyppi
- oppilaitoksen nimi
- oppilaitoksen sijaintipaikka
- vuosiluku

Viite ([Adida, 2006](#)) on esimerkki sähköisessä muodossa olevan opinnäytteen esittämisestä lähdeluettelossa.

Viite ([Kilpeläinen, 2001](#)) on esimerkki itsenäisen kirjoituksen sisältävästä verkkosivusta. Tällainen lähde on rinnastettavissa erillisteokseen. *Verkkosivusta* esitetään tiedot:

- tekijät
- otsikko
- maininta "Päivitetty" ja päivitysaika
- maininta "Viitattu" ja viittaamisen ajankohta
- Maininta "Saatavissa" ja URL.

Joskus verkkosivun kirjoitus on jaettu useammalle sivulle, jolloin lähdeluetteloon kirjataan vain sellainen verkko-osoite, joka koskee koko kirjoitusta tai sen etusivua, ellei sitten todella tarkoiteta kirjoituksen yksittäistä sivua.

2.7 Muuta huomioitavaa lähdeluettelossa

Lähdeluettelossa työn ja julkaisun nimi kirjoitetaan alkuperäisessä muodossaan. Julkaisijan kotipaikka kirjoitetaan alkukielisessä muodossaan.

Viittamista koskevassa suomalaisessa standardissa SFS 5342 ([Suomen standardisointiliitto, 2004](#)) vaaditaan julkaisuista ilmoitettavaksi myös ISBN- tai ISSN-numerot, mutta näissä opinnäyteohjeissa ei ISBN- ja ISSN-numeroita vaadita.

3 Tutkimusaineisto ja -menetelmät

Tässä osassa kuvataan käytetty tutkimusaineisto ja tutkimuksen metodologiset valinnat, sekä kerrotaan tutkimuksen toteutustapa ja käytetyt menetelmät.

4 Tulokset

Tässä osassa esitetään tulokset ja vastataan tutkielman alussa esitettyihin tutkimuskysymyksiin. Tieteellisen kirjoitelman arvo mitataan tässä osassa esitettyjen tulosten perusteella.

Tutkimustuloksien merkitystä on aina syytä arvioida ja tarkastella kriittisesti. Joskus tarkastelu voi olla tässä osassa, mutta se voidaan myös jättää viimeiseen osaan, jolloin viimeisen osan nimeksi tulee »Tarkastelu». Tutkimustulosten merkitystä voi arvioida myös »Johtopäätökset»-otsikon alla viimeisessä osassa.

Tässä osassa on syytä myös arvioida tutkimustulosten luotettavuutta. Jos tutkimustulosten merkitystä arvioidaan »Tarkastelu»-osassa, voi luotettavuuden arviointi olla myös siellä.

5 Yhteenveto

Opinnäytteen tekijä vastaa siitä, että opinnäyte on tässä dokumentissa ja opinnäytteen tekemistä käsittelevillä luennoilla sekä harjoituksissa annettujen ohjeiden mukainen muotoseikoiltaan, rakenteeltaan ja ulkoasultaan.

References

- B. Adida. Advances in cryptographic voting systems. *VTP Working Paper Series*, 51, 2006.
- J. Bardeen, L.N. Cooper, and J.R. Schrieffer. Theory of superconductivity. *Physical Review*, 108(5):1175–1204, 1957.
- G.A. Deschamps. Electromagnetics and differential forms. *Proceedings of the IEEE*, 69(6):676–696, 1981.
- I. Kauranen, M. Mustakallio, and V. Palmgren. Tutkimusraportin kirjoittamisen opas opinnäytetyön tekijöille, 2006.
- P Kilpeläinen. WWW-lähteisiin viittaaminen tutkielmatekstissä. *Verkkodokumentti*, 2001. Päivitetty 26.11.2001. Viitattu 19.1.2007. Saatavissa: <http://www.cs.uku.fi/~kilpelai/wwwlahteet.html>.
- N. Koblitz. *A course in number theory and cryptography*, volume 114. Springer Science & Business Media, 1994.

- S. Lindblom-Ylänne and M. Wager. Tieteellisten opinnäytetöiden ohjaaminen. In Lindblom-Ylänne S. and Nevgi A., editors, *Yliopisto- ja korkeakouluopettajan käsikirja*, pages 314–325. WSOY, Helsinki, 2002.
- A. Lönnqvist. *Applications of hologram-based compact range: antenna radiation pattern, radar cross section, and absorber reflectivity measurements*. Väitöskirja, Teknillinen korkeakoulu, Espoo, 2006.
- H. Miinusmaa. Neliskulmaisen reiän poraamisesta kolmikulmaisella poralla. Diplomityö, Teknillinen korkeakoulu, Espoo, 1977.
- V. Palmgren. Suunnittelija. Teknillinen korkeakoulu, kirjasto. Otaniementie 9, 02150 Espoo. Haastattelu 15.1., 2007.
- T. Pohjois-Koivisto. Voiko kone tulevaisuudessa arvata tahtosi? *Apropos*, Helmikuu 2005. Viitattu 19.1.2007. Saatavissa: <http://www.apropos.fi/1-2005/prima.php>.
- C.B. Ribeiro, E. Ollila, and V. Koivunen. Stochastic maximum-likelihood method for mimo propagation parameter estimation. *IEEE Transactions on Signal Processing*, 55(1):46–55, 2006.
- A. Sihvola, S. Tretyakov, P. Puska, and S. Kuehl. Interpretation of measurements of helix and bihelix superchiral structures. *Bianisotropics'98*, pages 317–320, 1998.
- T. Stieber. Gnupg hacks. *Linux Journal*, 143:2, 2006.
- Suomen standardisoimisliitto. SFS 5342, kirjallisuusviitteiden laatiminen, 2004.

A Esimerkki liitteestä

Liitteet eivät ole opinnäytteen kannalta välttämättömiä ja opinnäytteen tekijän on kirjoittamaan ryhtyessään hyvä ajatella pärjäävänsä ilman liitteitä. Kokemattomat kirjoittajat, jotka ovat huolissaan tekstiosan pituudesta, paisuttavat turhan helposti liitteitä pitääkseen tekstiosan pituuden annetuissa rajoissa. Tällä tavalla ei synny hyvää opinnäytettä.

Liite on itsenäinen kokonaisuus, vaikka se täydentääkin tekstiosaa. Liite ei siten ole pelkkä listaus, kuva tai taulukko, vaan liitteessä selitetään aina sisällön laatu ja tarkoitus.

Liitteeseen voi laittaa esimerkiksi listauksia. Alla on listausesimerkki tämän liitteen luomisesta.

```
\clearpage
\appendix
\addcontentsline{toc}{section}{Liite A}
\section*{Liite A}
...
\thispagestyle{empty}
...
tekstiä
...
\clearpage
```

Kaavojen numerointi muodostaa liitteissä oman kokonaisuutensa:

$$d \wedge A = F, \tag{A1}$$

$$d \wedge F = 0. \tag{A2}$$

B Toinen esimerkki liitteestä

Liitteissä voi myös olla kuvia, jotka eivät sovi leipätekstin joukkoon: Liitteiden



Figure B1: Kuvateksti, jossa on liitteen numerointi

taulukoiden numerointi on kuvien ja kaavojen kaltainen: Kaavojen numerointi

Table B1: Taulukon kuvateksti.

9.00–9.55	Käytettävyytestauksen tiedotustilaisuus (osanottajat ovat saaneet sähköpostitse valmistautumistehtävät, joten tiedotustilaisuus voidaan pitää lyhyenä).
9.55–10.00	Testausalueelle siirtyminen

muodostaa liitteissä oman kokonaisuutensa:

$$T_{ik} = -pg_{ik} + wu_i u_k + \tau_{ik}, \quad (\text{B1})$$

$$n_i = nu_i + v_i. \quad (\text{B2})$$