



Aalto-yliopisto
Perustieteiden
korkeakoulu

On comparison of software projects using various metrics with statistical analysis

Lauri Pulkkinen

06.03.2023

Ohjaaja: *VTM Teemu Kanerva*

Valvoja: *Assoc. Prof. Pauliina Ilmonen*

Työn saa tallentaa ja julkistaa Aalto-yliopiston avoimilla verkkosivuilla. Muilta osin kaikki oikeudet pidätetään.

Tausta

- Amerikkalaisen tutkimus- ja konsulttiyritys Standish Groupin vuosittaisissa raporteissa tulee ilmi, että yli puolet tietojärjestelmähankkeista myöhästyy tai budjetti ylittyy, tai laadussa ja toiminnallisuudessa on puutteita
- Vuonna 2019 julkaistun Standish Group raportin mukaan vain 16.2% ohjelmistoprojekteista oli onnistuneita, 52.7% projekteissa kustannus ja aikataulu ylittyivät ja/tai niistä puuttui luvattuja toiminnallisuuksia tai laadussa oli puutteita, 31.1% luokiteltiin hylätyiksi tai peruutetuiksi
- ”Väärä eli toimimaton tietojärjestelmä on yleistä julkishallinnossa”- Tomi Voutilainen
- ISO standardien tuntemus heikkoa
- Ohjelmistoprojektien mittaamisella pyritään ymmärtämään paremmin tämänhetkisten projektien tilaa ja ennustamaan tuotteen/prosessin tulevia ominaisuuksia
- Taustalla tarve löytää ongelmakohtia projektien etenemisessä ja löytää kohtia, missä on paranneltavaa, jotta tulevaisuuden projekteille olisi paremmat lähtökohdat
- Voidaan mitata useilla metriikoilla, esimerkiksi aikataulun varianssi, budjetin varianssi ja virhetiheys (defects/FP)
- Ohjelmiston mittaamisen prosessi on määritelty ja hallinnoitu ISO-standardeilla

Tavoitteet

- Tarkoituksena selvittää pysyvätkö Oikeusrekisterikeskuksen tarkasteluun valitut ohjelmistoprojektit budjetissa, aikataulussa ja onko ohjelmiston laadussa puutteita
- Ohjelmistoprojekteja vertaillaan tuottavuuden suhteen
- Mittausta varten valitaan sopivat metriikat, joita on maksimissaan kuusi kappaletta
- Projekteista kerätään tarvittava metriikka-aineisto ja lasketaan tarvittaessa johdettuja metriikoita ja epäsuoria metriikoita
- Käytetään tilastollisia menetelmiä, kuten tilastollista testausta, korrelaatioanalyysiä ja regressioanalyysiä
- Mallin testaaminen
- Lopuksi projekteja vertaillaan kerätyn aineiston perusteella
- Analyysiä voidaan käyttää tulevien projektien suunnittelussa, esim. miten suunnitellun kokoinen projekti on tyypillisesti edistynyt

Rajaukset

- Case-study, eli 4 projektia tarkasteluun
 - Tässä työssä pystytään selvittämään hyvin tämänhetkisten projektien tila ja saada vertailusta hyödyllistä tietoa, mutta useiden muuttujien vuoksi tulevien projektien ennustettavuus ei ole optimaalista
- Käytössä maksimissaan kuusi mittaria
 - Mittarien vertailukelpoisuus
- Aineiston saatavuus ja luotettavuus
 - Osa aineistosta saadaan valmiista tietokannoista, osa kerätään manuaalisesti
- Regressiomalli suuntaa antava
 - Pieni otoskoko vaikeuttaa tulkintaa ja saattaa luoda virheellistä päättelyä
 - Regressiomallin ennustuskykyä ei voida taata pienen otoskoon vuoksi

Tietolähteet ja aineistot

- Aineisto kerätään pääasiassa valmiista tietokannoista
- Osa aineistosta kerätään projekteista työtä varten
- Metriikoiden löytämiseen käytetään alan kirjallisuutta sekä internetiä
- Tilastollisten menetelmien valintaan ja soveltamiseen käytetään oppikirjoja ja tieteellisiä artikkeleita

Menetelmät ja työkalut

- Työn kirjoittamiseen käytetään LaTeX:ia
- Aineiston visualisointi ja analysointi toteutetaan yhdessä Excelillä sekä R:llä
- Osa visualisoinnista toteutetaan mahdollisesti Power BI:llä sekä PowerPointilla

Aikataulu

- Alan kirjallisuuteen, tietolähteisiin ja aineistoihin tutustuminen 01/2023-02/2023
- Aiheen esittely 6.3.2023
- Tekstin kirjoittaminen 02/2023-05/2023
- Aineiston kerääminen ja analysointi 03/2023-04/2023
- Valmiin työn esittely 16.5.2023