



**TURUN
YLIOPISTO**
Kauppakorkeakoulu

Pro gradu -tutkielma

Oppiaine: Tietojärjestelmätiede

Tekijä: Väinö Saarinen

Otsikko: Benefits from implementing a data labelling tool

Ohjaaja: Prof. Timo Leino

Sivumäärä: 71 sivua + 3 liitesivua

Päivämäärä: 6.6.2024

Datan määrä maailmassa kasvaa jatkuvasti ja sen hallinta on jatkuvasti yhä monimutkaisempaa ja vaativampaa. Tehokas datan hyödyntäminen on ratkaisevan tärkeää syvälliselle analyysille, loogiselle päättelylle ja päätöksenteolle. Datan luokittelu on olennainen osa tätä prosessia. Perinteisesti luokittelu on vaatinut merkittävästi työvoimaa ja resursseja. Jotta aina niukkoja resursseja voitaisiin kuluttaa tehokkaammin, yritykset kääntyvät datan luokittelutyökalujen puoleen automatisoidakseen prosessia, parantaakseen datan hallintaa ja luodakseen enemmän arvoa käytettävissä olevasta datasta.

Tämä tutkielma pyrkii arvioimaan datan luokittelutyökalun, erityisesti Microsoft Purview'n, käyttöönottoon liittyviä hyötyjä ja riskejä. Tutkimuksessa käytetään hyötyjen hallinta -mallia (benefit management model), ja tutkimukseen sisältyy yhteistyöyrityksessä toteutettu pilotti tutkimus. Lisävalidointia ja ammattimaisia näkemyksiä löydöksille tuovat yrityksen ammattilaisten haastattelut.

Tulokset paljastavat useita merkittäviä datan luokitteluun ja luokittelutyökaluihin liittyviä hyötyjä. Ensinnäkin luokittelutyökalut parantavat käytettävissä olevan datan laatua ja käyttäjän ymmärrystä käytettävissä olevasta datasta. Tämä lisää datan hyödyllisyyttä. Toiseksi automatisoidut luokittelutyökalut nopeuttavat huomattavasti luokitteluprosessia, vähentäen resurssien kulutusta verrattuna manuaalisiin menetelmiin. Kolmanneksi datan luokittelu tarjoaa monia etuja datan hallinnassa, tietojen hallinnoinnissa, tietojen katoamisen estämisessä, tietoturvassa, säännösten noudattamisessa ja datan elinkaaren hallinnassa. Datan luokittelutyökalun käyttöönottoon liittyvät riskit sisältävät riskit liittyen luokittelun tarkkuuteen, käyttäjien luokittelun omaksumiseen ja sitoutumiseen, sekä resurssien allokointiin.

Avainsanat: Datan luokittelutyökalu, Datan luokittelu, Hyötyjen hallinta