

EXAMENSARBETE Exploring subjectivity in ad hoc assessment of open source software**STUDENT** Jonathan Skogeby**HANDLEDARE** Martin Höst (LTH)**EXAMINATOR** Ulf Asklund (LTH)

Representera subjektiva kvalitetsbedömningar av GitHub-projekt med en linjärviktad modell

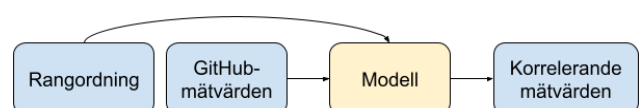
POPULÄRVETENSKAPLIG SAMMANFATTNING **Jonathan Skogeby**

Kvalitetsbedömning av projekt med öppen källkod sker ofta på improviserat vis med hög nivå av subjektivitet beroende på vem som utvärderar och vilka omständigheter projektet utvärderas för. I detta arbete har en enkel modell utvecklats och testats med syftet att försöka identifiera de mätvärden som använts under kvalitetsbedömningen.

GitHub är i särklass dagens mest använda plattform för att lagra kod och samarbeta på mjukvaruprojekt med andra utvecklare. Många användbara mätvärden kan överskådas på GitHub-sidan för ett projekt: Antalet bokmärkningar, följare, utvecklare och rapporterade buggar, senast tillagd kod, vilken licens som används, med mera. Oftast använder sig utvecklare av dessa mätvärden för att fastställa kvaliteten hos projektet, och jämför sedan med andra alternativ för att dra slutsatsen om vilket projekt som ska användas för att uppnå önskad funktionalitet.

Att välja fel mjukvarukomponent kan bli kostsamt. Bristande säkerhet, nedläggning av projektet och buggar utgör några av de risker som kommer med mjukvara med öppen källkod. Därför vill man se till att välja rätt redan från början. Dagens mjukvara innehåller ofta flera hundra komponenter med öppen källkod och att manuellt utvärdera alla alternativ är inte hållbart. För att utforska hur processen kan automatiseras måste det först avgöras vilka delar av processen som går att uttrycka i kod. I detta examensarbete har en modell utvecklats för att utforska möjligheten

att kartlägga vilka mätvärden på GitHub som använts av en viss utvecklare i sin kvalitetsbedömning. Baserat på data insamlad från GitHub och en utvecklares rangordning av en samling projekt kan den kombination av mätvärden som bäst korrelerar till rangordningen härledas av modellen. Följande figur ger en förenklad bild över modellens funktion.



Rangordningen av 20 projekt gjord av fem olika utvecklare samlades in tillsammans med mätvärdesdatan från samma projekt på GitHub. Resultatet pekade på att modellen ej lyckades fastställa vilka mätvärden som användes av utvecklarna när de rangordnade projekten trots att korrelationen mellan utvecklarnas och modellens rangordning var hög. Dock ger resultatet ej någon tydlig inblick i hur modellen hade presterat med en större datamängd.