

Datadriven Riskanalys ur ett DevOps-perspektiv

En populärvetenskaplig sammanfattning

Klimatet i dagens IT-sektor är sådant att man knappt kan bestämma sig för vilken kaffe man ska ha utan att först tillhandahålla data för att backa upp sitt beslut. Detta datadrivna paradigmskifte har slått rot i nästan alla delar av företagets aktiviteter och har fört med sig en uppsjö fördelar. Det finns dock ett område som vägrar låta sig datafieras - riskhantering. Till dags datum är riskhanteringsprocessen ofta löst baserad på allt ifrån subjektiv utvecklarcompetens till ren magkänsla. Denna artikel visar att denna process till viss del kan automatiseras, något som kan underlätta företags arbete oerhört.

För att kunna ta datadrivna beslut krävs data. Det är här ett annat nyligt paradigmskifte kommer in. **DevOps** är ett portmanteau av de engelska orden *Development* och *Operations*, och är ett arbetssätt som bl.a. lägger stor vikt vid att man sparar metainformation om utvecklingsprocessen. Ofta görs detta med hjälp av versioneringsverktyg såsom git och Jira. Dessa verktyg lagrar data om "commits", eller "ändringar". I vårt arbete var det ur denna datamängd de potentiella riskfaktorerna extraherades. Vår slutgiltiga lista över faktorer vi extraherade ur varje ändring är:

- Hur många olika filer som berörts
- Hur många rader som har lagts till
- Hur många rader som har tagits bort
- Tid sedan projektet ändrades senast
- Tiden på dagen då ändringen skedde
- Längden på meddelandet som beskrev ändringen



Med dessa faktorer extraherade var det dags för nästa steg. Riskhantering går ut på att minimera exponering till händelser med negativ påverkan. För att hitta det historiska sambandet mellan faktorerna och dessa negativa händelser utnyttjades en snillrik algoritm som går bakvägen från de ändringar som fixat kod. Tanken är att man utnyttjar dessa "fixes", för att hitta vilken kod som har behövts fixas. Sedan hittar man *ändringen* som gav upphov till dessa rader som behövde fixas - denna *ändring* blir sedan klassificerad som en källa till en negativ händelse.

Nu har vi en uppsättning riskfaktorer per ändring, och en approximering av om ändringen gav upphov till en negativ händelse eller inte. Allt som kvarstår att göra är att undersöka om det finns ett tillräckligt starkt samband för att kunna sia om ifall en kommande ändring medför stor eller liten risk. För att undersöka detta utnyttjades maskininlärning. Våra modeller visar att det finns ett tydligt samband mellan en *ändrings* faktorvärden och risk. Detta resultat kan användas som underlag för datadrivna beslut i samband med release av ny programvara och kod.
