

Hållbarhetsrapportering och ESG-mått med hjälp av maskininlärning

Peter Öhman och Jan Svanberg



Hållbarhetsrapportering och ESG-mått med hjälp av maskininlärning

Peter Öhman och Jan Svanberg

Tidigare studier visar att kvaliteten på olika hållbarhetsmått kan ifrågasättas. Det gäller inte minst så kallade ESG-mått (där ESG står för environmental, social och governance). Ett problem är att kreditvärderingsföretag tenderar att lägga samman olika hållbarhetsindikatorer utan att ta hänsyn till hur väsentliga de är. I stället används aritmetiska medelvärden för att kunna jämföra varje företag med ett genomsnittligt företag. Mot den bakgrunden har Peter Öhman och Jan Svanberg vid CER, tillsammans med kollegor från andra universitet i och utanför Sverige, utvecklat en ny metod för att beräkna dessa mått. Den föreslagna metoden bygger på maskininlärningsteknik för att aggregera indikatorer och med hjälp av de mönster som då framträder kunna förutsäga sannolikheten för enskilda företags inblandning i så kallade ESG-kontroverser. Forskningsresultaten har tidigare publicerats i tre internationella tidskriftsartiklar med fokus på var och en av de tre hållbarhetsaspekterna E, S respektive G: "Prediction of environmental controversies and development of a corporate environmental performance rating methodology", "Must social performance ratings be idiosyncratic? An exploration of social performance ratings with predictive validity" och "Corporate governance performance ratings with machine learning" (Svanberg et al., 2022a; 2023a; 2022b).

Inledning

Trots utbredningen av hållbarhetsfrågor runt om i världen saknas tillförlitliga metoder för att mäta och värdera nivån på hållbara prestationer hos enskilda företag.¹ De flitigast använda måtten – så kallade ESG-mått, där de tre bokstäverna står för environmental, social och governance² – aggregerar olika hållbarhetsindikatorer med hjälp av aritmetiska medelvärden i syfte att göra jämförelser med ett genomsnittligt företag. De ESG-betyg som olika kreditvärderingsföretag sedan tilldelar enskilda företag har å ena sidan blivit populära, men innehåller å andra sidan flera allvarliga brister.

För det första visar tidigare studier att hållbarhetsbegreppet är både omfattande och heterogent samt att de konventionella ESG-betygen förvisso innehåller väsentliga indikatorer, men att de förenklade beräkningsmetoder som används för att aggregera holistiska betyg inte gör rättvisa åt de olika indikatorernas vikt och betydelse.³ Hur väsentliga är till exempel var och en av tre så olika ESG-indikatorer som hantering av farligt avfall, åsidosättande av personalens rättigheter och bristande företagsstyrning? Antagandet att alla ESG-aspekter och alla indikatorer är lika viktiga för företags

¹ Berg et al. (2022).

² Environmental avser miljö, social avser socialt ansvar och governance avser bolagsstyrning.

³ Chatterji et al. (2016).

hållbarhet stämmer helt enkelt inte. Mångfacetterade områden kräver förhållandevis komplexa mätverktyg.⁴

För det andra skiljer sig bedömningarna från olika kreditvärderingsföretag åt högst väsentligt. Studier visar att ESG-betygen kan skilja sig så mycket som 30 procent av betygssiffran när olika kreditvärderingsföretag bedömer samma företags hållbarhetsinsatser.⁵

För det tredje innehåller inte nuvarande ESG-betyg den information som institutionella investerare behöver för att bedöma om ett företag är mer hållbart än ett annat och utifrån det kunna sätta samman investeringsportföljer som tar hänsyn till reella ESG-risker.⁶ Aritmetiska medelvärden är mer eller mindre irrelevanta som grund för investeringsbeslut om huruvida företag A eller företag B förtjänar (mest) beröm från ett hållbarhetsperspektiv. Ju fler ESG-indikatorer en investeringsbedömning omfattar, desto svårare blir arbetet för institutionella investerare eftersom investeringsbeslutet i slutändan bygger på allt för många och allt för olika indikatorer för att kunna slås samman till rättvisande helhetsbetyg.

För det fjärde relateras inte konventionell aggregering av ESG-indikatorer till någon rättslig eller annan jämförbar standard.⁷ Domstolar bedömer individers beteenden i förhållande till gällande lagar och utan en sådan referenspunkt skulle bedömningarna vara godtyckliga, tvivelaktiga och betraktas som tämligen meningslösa. Ett annat exempel är att revisorer bedömer informationen i företags finansiella rapporter i förhållande till lagar, rekommendationer och redovisningsstandarder. Avsaknaden av en referenspunkt i fråga om hållbarhet lämnar en besvärande fråga obesvarad: Vad mäter de oförankrade aggregerade ESG-betygen egentligen?

Syfte och tillvägagångssätt

För att komma till rätta med de nuvarande ESG-betygens tillkortakommanden kan maskininlärning användas för att utveckla en ny typ av metodologi med fokus på i vilken utsträckning enskilda företag följer de delar av hållbarhetsarbetet som kan utlösa kontroverser och som av det skälet kan vara av intresse för alla intressentgrupper.

Kontroverser kan beskrivas som reaktioner på företags hållbarhetsarbete, exempelvis hantering av avfall. Denna typ av information är artskild från den som vanligtvis summeras till ESG-betyg. Hållbarhetskontroverser säger också något om i vilken mån

⁴ Wood (2010)

⁵ Chatterji et al. (2016); Christensen et al. (2022).

⁶ Nofsinger et al. (2019).

⁷ Svanberg et al. (2023b).

företag har ändamålsenliga strukturer och processer eftersom kontroverser ofta uppstår till följd av medvetna handlingar.

Vad som är särskilt intressant för institutionella investerare är företags efterlevnad av bindande beteendestandarder eftersom den typen av försummelser riskerar att orsaka oåterkallelig ekonomisk skada.⁸ Dessa standarder kan vara juridiskt bindande, (till exempel att agera enligt arbets-, miljö- och bolagsrätten), följa så kallad "soft law" (som att konsekvent hålla sig till gällande internationella redovisningsrekommendationer) samt innefatta moraliska förpliktelser (till exempel att ta klimatförändringar på allvar).

Iakttagelsen att denna typ av försummelser riskerar att orsaka ekonomisk skada stärks av att aktiemarknaden visar små eller inga reaktioner på goda nyheter men stora negativa reaktioner på dåliga nyheter.⁹ För att skydda sitt eget rykte gör varje institutionell investerare därför klokt i att undvika investeringar i företag som är inblandade i så kallade ESG-kontroverser.¹⁰ Negativa ekonomiska effekter kan bland annat vara förknippade med kontroverser som uppstår i form av rättstvister initierade av statliga eller internationella organisationer, bojkotter och strejker initierade av olika intressentgrupper eller misstänkta oegentligheter respektive djupt oetiska beteenden uppmärksammas av massmedier.

Syftet med de studier som ligger till grund för den här kortrapporten är att beskriva och diskutera en maskininlärningsmetodik med vilken ESG-betyg med hög precision kan genereras. Baserat på tidigare studier är ambitionen att metoden ska ha prediktiv förmåga och att den ska fokusera på kontroverser kopplade till enskilda företag.¹¹

De tre studierna, en per hållbarhetsaspekt, utgår från nio maskininlärningsmetoder med delvis olika funktioner och egenskaper för att kartlägga mönster av olika indikatorer och med det som grund förutsäga risken för ESG-kontroverser i specifika företag eller företag i en och samma bransch. Utgångspunkten är nämligen att indikatorsmönster kan förknippas med ett företags risk att uppleva en kontrovers. Med vårt datadrivna tillvägagångssätt undersöks samband mellan indikatorer hämtade från hela hållbarhetsområdet och kontroverser beskrivna i massmedierna under åren 2009–2018.

Datan är hämtad från internationella företagsdatabaser och består av 400 indikatorer (som alltså täcker in de tre hållbarhetsaspekterna E, S och G). Studierna baseras vidare på ett urval av 2517 företag i olika branscher och i olika länder för den undersökta tioårsperioden. Företagen delades in i två grupper: en bestående av kontrovers-

⁸ Nofsinger et al. (2019).

⁹ Capelle-Blancard och Petit (2019).

¹⁰ Krueger et al. (2020).

¹¹ Ding et al. (2020).

benägna företag, det vill säga de som har upplevt E- respektive G-kontroverser eller som har upplevt fler än ett genomsnittligt antal S-kontroverser, och den andra bestående av företag som inte har upplevt några E- respektive G-kontroverser eller har upplevt färre S-kontroverser än genomsnittet. Skillnaden i hur E- och G- respektive S-kontroverser behandlas beror på att den sistnämnda typen av kontroverser är mer förekommande än övriga två.

Resultat

Tabell 1 visar med vilken precision två av maskininlärningsmetoderna – Random forest, som är en av de mest komplexa och samtidigt den mest kapabla metoden, och Logistisk regression, som är mindre komplex och mindre kapabel men fortfarande relativt välpresterande – kan förutsäga risken för att ett företag får uppleva var och en av de tre typerna av kontroverser (E, S och G). Precisionsmåttet är särskilt intressant för institutionella investerare eftersom de ska allokera stora summor pengar utifrån hur hållbart vart och ett av företagen i en investeringsportfölj är.

Som framgår av tabellen visar Random forest genomgående höga siffror: 70–84 procent av kontroverserna är korrekt förutspådda. Med Logistisk regression kan mellan drygt 30 och drygt 60 procent av kontroverserna förutspås. Båda metoderna har högst precision vad gäller S-kontroverser.

Tabell 1: Precisionsgrad i procent för två av maskininlärningsmetoderna avseende de tre typerna av kontroverser (E, S respektive G)

	Precision
Environmental	
Random forest	69,94
Logistisk regression	30,45
Social	
Random forest	84,17
Logistisk regression	61,32
Governance	
Random forest	75,06
Logistisk regression	33,87

Utifrån möjligheten att utforma en prediktiv modellering av ESG-kontroverser förväntar vi oss att alla ESG-betyg, i synnerhet de betyg som grundar sig på denna

modellering, ska vara negativt korrelerade med det antal kontroverser som ett företag upplever. Genom att de AI-genererade metoderna studerar samband mellan mönster av hållbarhetsindikatorer och kontroverser är tanken att de mönster som är förknippade med kontroverser ska leda till att företag som upplever jämförelsevis fler kontroverser tilldelas lägre betyg än andra företag.

Det kan dock finnas avvikelser från det normala. Till exempel kan företag som upplever fler ESG-kontroverser än genomsnittet ha indikatorsmönster som är typiska för företag som "ser bra ut" ur metodens synvinkel. Andra företag kan ha hamnat i inga eller färre kontroverser än genomsnittet trots att de uppvisar ett mönster av egenskaper som är typiska för företag som är utsatta för kontroverser. Metoden kan således inte med säkerhet förutsäga att ett företag kommer att hamna i en eller flera kontroverser, utan enbart att det finns en risk för att företag med ett visst indikatorsmönster ska uppleva en eller flera mer eller mindre allvarliga hållbarhetskontroverser.

Slutdiskussion

Vikten av ansvarsfulla investeringar stöds av ett stort antal institutionella investerare i de flesta delar av världen. Dessa investerare har möjlighet att påverka företag att bli mer hållbara genom att utöva sin ägarmakt. Det är dock svårt för institutionella investerare att i god tro investera ansvarsfullt och påverka företagen i investeringsportföljerna att bli mer hållbara om de inte vet vad som kännetecknar ett hållbart företag. Konventionell metodik för aggregering av ESG-indikatorer bidrar inte mer än marginellt till hållbara investeringar eftersom det är större sannolikhet att sådana metoder mäter kapaciteten för stora företags hållbarhetsretorik än deras hållbara prestationer.¹²

En datadriven lösning på problemet är att bedöma företags ageranden i förhållande till en prestationsstandard och vi förankrar vår metod i ESG-kontroverser, det vill säga en proxy för företags efterlevnad inom hållbarhetsområdet.

Användningen av ESG-kontroverser förankrar både hållbarhetsindikatorer och företags ESG-betyg i ett system av bindande beteendestandarder. Företag som inte följer dessa riskerar per definition att göra sig till ett mål för kontroverser. Samtidigt finns det goda skäl att tro att företag som är inblandade i kontroverser kännetecknas av att de inte har ändamålsenliga strukturer och processer på plats eller att den etiska kompassen inte är tillräckligt stark. Ett ESG-betyg grundat på kontroverser skulle således mäta de mest grundläggande förutsättningarna för företags legitimitet och vara förenligt med institutionella investerares informationsbehov.

¹² Dremptic et al. (2020).

Maskininlärningsmetoderna lär sig att bedöma vikterna på eller betydelsen av olika hållbarhetsindikatorer. Den bästa av de nio använda metoderna (Random forest) producerar inte bara ett viktningsschema som speglar företagets förmåga att leva upp till bindande beteendestandarder, den komplexa och icke-linjära metoden ger även exempel på den mångfacetterade interaktion som kan förekomma mellan ESG-indikatorer. Genom att extrapolera relationerna mellan indikatorsmönster och kontroverser för alla företag kan maskininlärningsmetoderna förutsäga risken för att ett specifikt företag brister i efterlevnaden av bindande beteendestandarder (förutsatt att det finns tillgång till relevant indata). Den bästa av våra metoder har hög prediktiv validitet och tillgodoser det informationsbehov som institutionella investerare har.

Noteras ska att den pågående metodutvecklingen ännu bara är i sin linda och att flera begränsningar måste övervinnas. Framtida metoder skulle exempelvis kunna använda longitudinella data i syfte att förutsäga risken för framtida kontroverser. Metoderna tar heller inte hänsyn till att kontroverser inom ramen för de tre hållbarhetsaspekterna E, S respektive G kan skilja sig från varandra högst väsentligt och att olika ESG-indikatorer kan ha olika associationer till olika kontroverser. Hänsyn behöver också tas till att företagsstorlek och mediauppmärksamhet sannolikt är positivt korrelerade och att det resulterar i att ESG-betygen tenderar att bli lika mycket storleksorienterade som kontroversorienterade. En rekommendation är därför att inrikta den fortsatta metodutvecklingen på sådana missförhållanden som signaler kontroverser snarare än det mediala intresse som stora företag per automatik har att leva med.¹³

Trots olika begränsningar tyder resultaten på att de AI-genererade ESG-betygen har större relevans än de konventionella betyg som bygger på att ESG-indikatorer läggs samman med hjälp av aritmetiska medelvärden. Maskininlärningstekniker är således ett hoppgivande steg på vägen från förenklad till prediktiv modellering. Vidareutvecklade maskininlärningsmetoder har också förutsättningar att bidra till att ESG-betygen blir ett både objektiva och transparent mått för institutionella investerare att utgå ifrån.

¹³ Fiaschi et al. (2020)

Referenser

- Berg, F., Kölbel, J.F. och Rigobon, R. (2022), "Aggregate confusion: The divergence of ESG ratings". *Review of Finance*, Vol. 26, Nr. 6, s. 1315-1344.
- Capelle-Blancard, G. och Petit, A. (2019), "Every little helps? ESG news and stock market reaction". *Journal of Business Ethics*, Vol. 157, Nr. 2, s. 543-565.
- Chatterji, A.K., Durand, R., Levine, D.I. och Touboul, S. (2016), "Do ratings of firms converge? Implications for managers, investors and strategy researchers". *Strategic Management Journal*, Vol. 37, Nr. 8, s. 1597-1614.
- Christensen, D., Serafeim, G. och Sikochi, A. (2022), "Why is corporate virtue in the eye of the beholder?". *The Accounting Review*, Vol. 97, Nr. 1, s. 147-175.
- Drempetic, S., Klein, C. och Zwergel, B. (2020), "The influence of firm size on the ESG score: Corporate sustainability ratings under review". *Journal of Business Ethics*, Vol. 167, Nr. 1, s. 333-360.
- Ding, K., Lev, B., Peng, X., Sun, T. och Vasarhelyi, M.A. (2020), "Machine learning improves accounting estimates: Evidence from insurance payments". *Review of Accounting Studies*, Vol. 25, Nr. 3, s. 1098-1134.
- Fiaschi, D., Giuliani, E., Nieri, F. och Salvati, N. (2020), "How bad is your company? Measuring corporate wrongdoing beyond the magic of ESG metrics". *Business Horizons*, Vol. 63, Nr. 3, s. 287-299.
- Krueger, P., Sautner, Z. och Starks, L. (2020), "Importance of climate risks for institutional investors". *Review of Financial Studies*, Vol. 33, Nr. 3, s. 1067-1111.
- Nofsinger, J.R., Sulaeman, J. och Varma, A. (2019), "Institutional investors and corporate social Responsibility". *Journal of Corporate Finance*, Vol. 58, Nr. C, s. 700-725.
- Svanberg, J., Ardeshiri, T., Samsten, I., Öhman, P., Rana, T. och Danielson, M. (2022a), "Prediction of environmental controversies and development of a corporate environmental performance rating methodology". *Journal of Cleaner Production*, Vol. 344, 130979.
- Svanberg, J., Ardeshiri, T., Samsten, I., Öhman, P., Neidermeyer, P., Rana, T., Semenova, N. och Danielson, M. (2022b), "Corporate governance performance ratings with machine learning". *Intelligent Systems in Accounting, Finance and Management*, Vol. 29, Nr. 1, s. 50-68.
- Svanberg, J., Ardeshiri, T., Samsten, I., Öhman, P., Neidermeyer, P., Rana, T., Maisano, F. och Danielson, M. (2023a), "Must social performance ratings be idiosyncratic? An exploration of social performance ratings with predictive validity". *Sustainability Accounting, Management and Policy Journal*, Vol. 14, Nr. 7, s. 313-348.
- Svanberg, J., Ardeshiri, T., Samsten, I., Öhman, P. och Neidermeyer, P. (2023b), "Prediction of controversies and estimation of ESG performance: An experimental investigation using machine learning". I Rana, T., Svanberg, J., Öhman, P. och Lowe, A. (Red.), *Handbook of big data and analytics in accounting and auditing*, Singapore: Springer, s. 65-88.
- Wood, D.J. (2010), "Measuring corporate social performance: A review". *International Journal of Managerial Reviews*, Vol. 12, Nr. 1, s. 50-84.