

Skript

Erarbeitet von

Ludmila Himmelpach und Gerhard Klassen

Lernziele	1
Inhalt	1
Einstieg.....	1
Betrugserkennung in Finanzdaten	2
Herausforderungen bei der Erkennung von Bilanzfälschung.....	2
Zeitliches Clustering von Unternehmensdaten.....	3
Ausreißer als mögliche Betrugsfälle.....	4
Take-Home Message	5
Quellen	5
Disclaimer	6

Lernziele

- Du erfährst über die Herausforderungen der automatisierten Erkennung von Bilanzfälschung.
- Du lernst, wie Clusteringmethoden für die automatisierte Erkennung von möglichen Bilanzfälschungen eingesetzt werden können.

Inhalt

Einstieg

Du weißt bereits, dass Clusteringalgorithmen dazu benutzt werden können, um Kunden in Gruppen zu segmentieren, die ähnliche Interessen oder ähnliches Verhalten aufweisen. In einem Gastbeitrag von Dr. Gerhard Klassen erfährst du, wie Clusteringmethoden zur Aufdeckung von potentiellen Bilanzfälschungen beitragen können.

Betrugserkennung in Finanzdaten

Börsennotierte Unternehmen sind gesetzlich verpflichtet, ihre Finanzen jährlich in Form einer Bilanz offenzulegen. Doch nicht alle Unternehmen nehmen es dabei ganz genau mit der Wahrheit. Manche Wenige manipulieren bewusst ihre Zahlen – und das mit klaren Absichten: etwa um Investoren und Anleger zu täuschen oder ihre Kreditwürdigkeit besser darzustellen, um leichter an Kredite zu kommen.

Immer wieder hört man aus den Medien von entsprechenden Fällen, wie zuletzt beispielsweise von Wirecard, Theranos oder 2001 Enron – einem der größten Betrugsfälle des letzten Jahrhunderts.

Quelle [1,2,3]

Die Bilanzfälschung ist strafbar und ihre Aufdeckung ist insofern wichtig als dass eine bewusste Falschangabe in Jahresabschlüssen eine große Auswirkung auf Anleger, Investoren und den gesamten Markt hat und folglich einen großen Schaden anrichten kann.

Herausforderungen bei der Erkennung von Bilanzfälschung

Wer gefälschte Bilanzen aufdecken möchte, steht vor mehreren Herausforderungen. Denn neben absichtlichen Manipulationen gibt es auch ganz banale Fehler – zum Beispiel Zahlendreher oder Übertragungsfehler. Wenn solche Unstimmigkeiten auffallen, müssen Unternehmen ihre Bilanz korrigieren. Das passiert in etwa 2 % aller jährlich eingereichten Bilanzen.

Ein weiteres Problem: Nur wenige Betrugsfälle sind überhaupt bekannt. Es ist jedoch sehr wahrscheinlich, dass viele Manipulationen unentdeckt bleiben. Und das vielleicht größte Problem: Betrug sieht jedes Mal anders aus. Es gibt kein einheitliches Muster – schließlich fällt kaum jemand zweimal auf dieselbe Masche rein.

Genau hier stoßen überwachte Lernverfahren, wie Klassifikationsalgorithmen, an ihre Grenzen. Die Daten wären stark unausgeglichen, und die Fälle lassen sich nicht eindeutig voneinander abgrenzen – das macht es für ein Modell schwierig, verlässliche Muster zu erkennen.

Deshalb nähern wir uns dem Problem mit Methoden des unüberwachten Lernens. Konkret suchen wir nach Unternehmen, die sich im Vergleich zu ähnlichen Unternehmen auffällig verhalten – also sogenannte Ausreißer.

Quelle [4]

Zeitliches Clustering von Unternehmensdaten

Betrachten wir zunächst eine einfache Abstraktion des Problems. Hier sehen wir sieben Zeitreihen von sieben Unternehmen – sie zeigen die Entwicklung eines beliebigen Bilanzwerts über die Zeit.

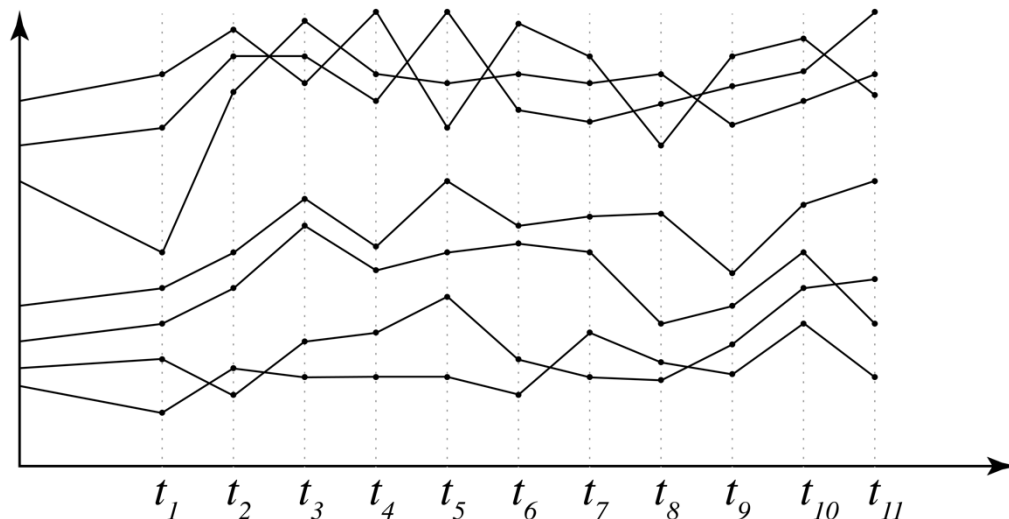


Abbildung 1: Eine beispielhafte Darstellung der Bilanzdaten von sieben Unternehmen zu verschiedenen Zeitpunkten.

Wenn man sich eine einzelne Zeitreihe isoliert anschaut, ist es schwer zu beurteilen, ob sie auffällig ist – denn es fehlt der Kontext. Die finanzielle Entwicklung eines Unternehmens hängt oft stark von äußeren Faktoren ab – wie wirtschaftlichen Krisen oder dem Boom einzelner Branchen.

Deshalb ist es sinnvoll, viele Unternehmen gemeinsam zu betrachten – und gezielt nach Abweichungen im Vergleich zu ähnlichen Unternehmen zu suchen.

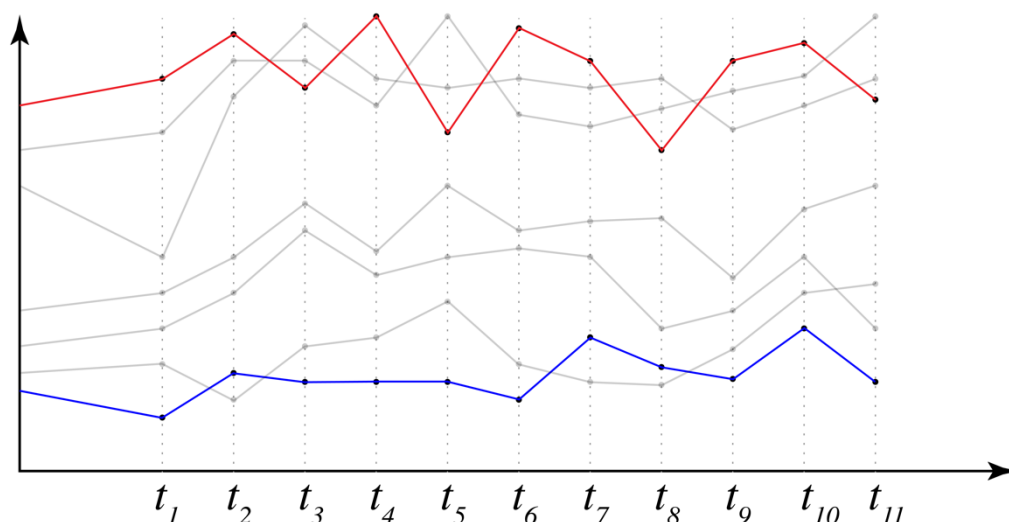


Abbildung 2: : Eine beispielhafte Darstellung der Bilanzdaten von sieben Unternehmen zu verschiedenen Zeitpunkten.

Da sich ein Energieversorger vermutlich ganz anders entwickelt als ein Getränkehersteller, wollen wir zunächst herausfinden, welche Unternehmen sich in ihrem Verhalten ähneln. Dafür clustern wir die Kennzahlen aller Unternehmen zu jedem Zeitpunkt. So wird sichtbar, dass sich manche Unternehmensverläufe über die Zeit hinweg sehr ähnlich sind – zum Beispiel die Unternehmen, die hier dem grünen Cluster zugeordnet wurden.

Quelle [5, 6]

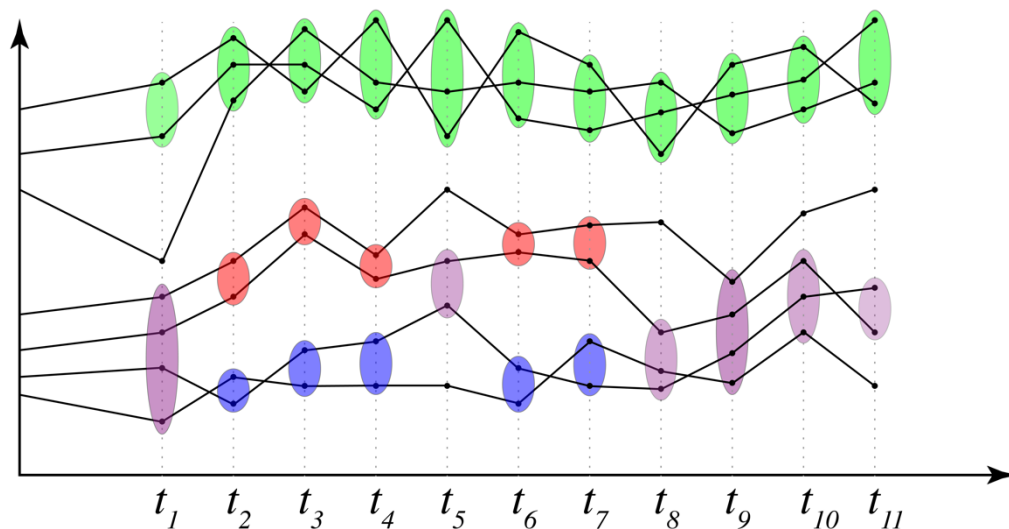


Abbildung 3: Eine beispielhafte Darstellung der Bilanzdaten von sieben Unternehmen zu verschiedenen Zeitpunkten nach dem Clustering.

Andere Unternehmen wechseln hingegen im Laufe der Zeit ihre Clusterzugehörigkeit – oder bilden sogar neue Cluster, wenn sie sich stark von den anderen unterscheiden. Manche lassen sich zeitweise gar keinem Cluster zuordnen.

Aber was genau sind hier nun Ausreißer?

Ausreißer als mögliche Betrugsfälle

Im klassischen Clustering sind das jene Datenpunkte, die keinem Cluster zugeordnet werden konnten. In unserem Fall ist jedoch die zeitliche Entwicklung entscheidend: Ausreißer sind die Zeitreihen, die ihre Clusternachbarn besonders häufig wechseln.

Die auffälligste Zeitreihe in unserem Beispiel ist die vierte von oben – sie wechselt ihre Clusterzugehörigkeit häufiger als alle anderen.

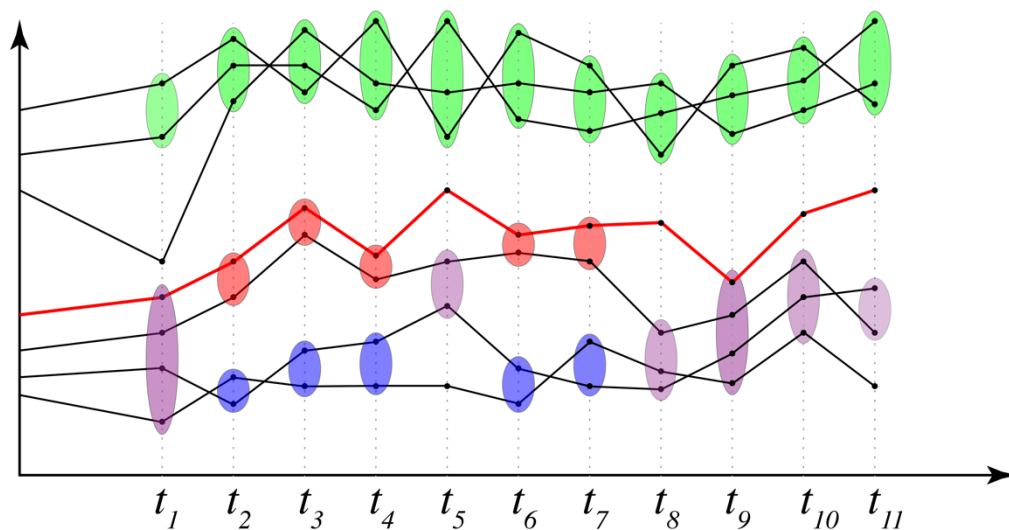


Abbildung 4: Eine beispielhafte Darstellung der Bilanzdaten von sieben Unternehmen zu verschiedenen Zeitpunkten nach dem Clustering.

Wenn sich ein Unternehmen trotz Ähnlichkeit zu anderen plötzlich auffällig entwickelt, lohnt sich ein genauerer Blick. In unserer Untersuchung konnten wir mit dieser Methode einen Großteil der Bilanzberichtigungen vorab identifizieren – also genau jene Fälle, in denen die Zahlen später offiziell korrigiert werden mussten. In einigen Fällen wurden diese Korrekturen nicht durchgeführt – vielleicht sind wir hier etwas auf der Spur :)

Quelle [7]

Take-Home Message

In diesem Video hast du erfahren, welchen Herausforderungen man bei der automatisierten Erkennung von Bilanzfälschung begegnet, und wie Clusteringmethoden zur Aufdeckung von potentiellen Betrugsfällen beitragen können.

Quellen

- Quelle [1] Wischmeyer, N. (2023, 13. Februar). Aufstieg und Fall von Wirecard. *Süddeutsche Zeitung*. <https://www.sueddeutsche.de/wirtschaft/wirecard-zusammenfassung-aufstieg-und-fall-1.5750529>
- Quelle [2] Gleim, F. (2022, 11. März). Enron: Ein Meisterwerk der Bilanzfälschung. *BANKINGCLUB*. <https://www.bankingclub.de/news/anlage/enron-ein-meisterwerk-der-bilanzfaelschung/>
- Quelle [3] Lindner, R. (2023, 30. Mai). Elizabeth Holmes tritt Gefängnisstrafe an. *Frankfurter Allgemeine Zeitung*. <https://www.faz.net/aktuell/wirtschaft/unternehmen/theranos-elizabeth-holmes-tritt-gefaengnisstrafe-an-18929498.html>

- Quelle [4] Klassen, G., Tatusch, M., Huo, W., & Conrad, S. (2020, August). Evaluating machine learning algorithms in predicting financial restatements. In *2020 The 4th International Conference on Business and Information Management* (pp. 95-100).
- Quelle [5] Tatusch, M., Klassen, G., Bravidor, M., & Conrad, S. (2020). How is your team spirit? Cluster over-time stability evaluation. In *16th international conference on machine learning and data mining, MLDM* (pp. 155-170).
- Quelle [6] Klassen, G., Tatusch, M., Himmelspace, L., & Conrad, S. (2020, June). Fuzzy clustering stability evaluation of time series. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems* (pp. 680-692). Springer International Publishing.
- Quelle [7] Tatusch, M., Klassen, G., Bravidor, M., & Conrad, S. (2019, November). Show Me Your Friends and I'll Tell You Who You Are. Finding Anomalous Time Series by Conspicuous Cluster Transitions. In *Australasian Conference on Data Mining* (pp. 91-103). Singapore: Springer Singapore.

Disclaimer

Transkript zu dem Video „Woche 09 Anwendung: Anwendung von Clustering zur Erkennung von Bilanzfälschung“, Ludmila Himmelspace und Gerhard Klassen.

Dieses Transkript wurde im Rahmen des Projekts ai4all des Heine Center for Artificial Intelligence and Data Science (HeiCAD) an der Heinrich-Heine-Universität Düsseldorf unter der Creative Commons Lizenz [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/) veröffentlicht. Ausgenommen von der Lizenz sind die verwendeten Logos, alle in den Quellen ausgewiesenen Fremdmaterialien sowie alle als Quellen gekennzeichneten Elemente.