

Bachelorarbeit  
Features of Business Success

Matthias Wagner  
e1227617@student.tuwien.ac.at  
Matrikelnummer: 1227617  
Studienkennzahl: 033 532

Betreuer: Ao.Univ.Prof. Mag. Dr. Horst Eidenberger

Mai 2015

# 1 Projektbeschreibung

## 1.1 Aufgabenstellung

Ziel der Arbeit ist es die Features zu identifizieren, die Erfolg in der Wirtschaft zugrunde liegen. Dazu soll eine Testdatenbank angelegt und mit Bilanzdaten österreichischer Unternehmen, verbunden mit einer Erfolgs-bewertenden Ground-Truth, befüllt werden. Auf diese Daten soll anschließend

- 1.) ein Feature-Extraktions-Verfahren angewandt werden, um aussagekräftige Kennzahlen zu identifizieren und
- 2.) versucht werden mit Hilfe eines Machine-Learning-Systems zukünftigen Börsenerfolg, anhand vorhandener Kennzahlen, vorherzusagen.

## 1.2 Zeitliste

Datum	von	bis	Zeit	Tätigkeit
26.02.15	10:00	10:30	00:30	Vorbesprechung
06.03.15	13:00	13:30	00:30	Planen; Meilensteinplan
08.03.15	13:30	15:30	02:00	Recherche - Suchen von Daten
09.03.15	10:30	12:00	01:30	Recherche - Suchen von Daten
10.03.15	12:00	14:00	02:00	Recherche; Daten sammeln
11.03.15	10:30	14:30	04:00	Recherche; Kennzahlen selektieren; Datensätze erstellen
12.03.15	10:30	12:30	02:00	Datensätze erstellen
13.03.15	09:30	11:30	02:00	Datensätze erstellen
13.03.15	15:00	16:00	01:00	Datensätze erstellen
14.03.15	10:30	12:00	01:30	Datensätze erstellen
15.03.15	11:00	13:30	02:30	Datensätze erstellen
18.03.15	11:00	12:00	01:00	Datensätze überarbeiten; Börsenkurse recherchieren
18.03.15	13:45	15:30	01:45	Börsenkurse recherchieren
20.03.15	09:30	11:30	02:00	Börsenkurse recherchieren
20.03.15	15:30	16:45	01:15	Dokumentation der Datensätze
21.03.15	10:00	11:00	01:00	Daten: Arff Datei erstellen; Java-Projekt aufsetzen
27.03.15	10:30	12:00	01:30	Weka lernen
16.04.15	17:15	20:15	03:00	Abschlussbericht strukturieren; mit WEKA experimentieren
17.04.15	12:45	14:15	01:30	Abschlussbericht strukturieren; ergänzen
22.04.15	15:15	17:30	02:15	Ergebnisse auswerten; Abschlussbericht schreiben
23.04.15	15:45	19:30	03:45	Abschlussbericht schreiben; Analyse mit R
22.05.15	14:30	16:00	01:30	Abschlussbericht schreiben
24.05.15	12:30	13:00	00:30	Korrekturlesen
27.05.15	17:30	18:00	00:30	Finalisieren

Gesamt: 41:00

# 2 Implementierung

## 2.1 Technische Details

### Systemumgebung

Mac OSX Yosemite, Version 10.10.2

### Verwendete Bibliotheken und Programme

Weka 3.6.12

RStudio 0.98.1103 mit R 3.2.0

## 2.2 Vorgehensweise

1. Testdaten erstellen
2. Analyse mit Weka und R:
  - (a) Feature Extraktion
  - (b) Erfolg vorhersagen

## 2.3 Testdaten

Insgesamt gibt es 100 Datensätze, im ARFF Format, von 50 österreichischen Aktiengesellschaften. Die dafür verwendeten Daten entstammen den Bilanzen der jeweiligen Jahresfinanzberichte, aus dem Zeitraum 31.12.2011 bis 31.3.2014, die auf der Internetseite der Österreichischen Kontrollbank erhältlich sind.[1]

Da sich Bank-Bilanzen zum Teil stark von Unternehmens-Bilanzen unterscheiden werden keine Daten von Banken verwendet.

Die Datensätze bestehen aus dem jeweiligen Firmennamen, Abschlussdatum der Bilanz, der Ground-Truth, die "Erfolg" oder "kein Erfolg" symbolisiert, und folgenden 23 betriebswirtschaftlichen Kennzahlen:

- Direkt der Bilanz entnommene Kennzahlen: (in TEUR)
  - Gesamtkapital
  - Anlagevermögen
  - Umlaufvermögen
  - Eigenkapital
  - Grundkapital
  - Rückstellungen
  - Verbindlichkeiten
  - Umsatzerlöse
  - Betriebserfolg
  - Finanzerfolg
  - EBT
  - Jahresüberschuss
- Aus der Bilanz berechnete Kennzahlen:
  - Eigenkapitalrentabilität ( $=\text{Jahresüberschuss}/\text{Eigenkapital}$ )
  - Gesamtkapitalrentabilität ( $=\text{EBT}/\text{Gesamtkapital}$ )
  - Umsatzrentabilität ( $=\text{Betriebserfolg}/\text{Umsatz}$ )
  - Kapitalumschlag ( $=\text{Umsatz}/\text{Gesamtkapital}$ )
  - Deckungsgrad A ( $=\text{Eigenkapital}/\text{Anlagevermögen}$ )

- Vermögensintensität (=Anlagevermögen/Umlaufvermögen)
- Anlageintensität (=Anlagevermögen/Gesamtkapital)
- Umlaufintensität (=Umlaufvermögen/Gesamtkapital)
- Sachanlagenbindung (=Anlagevermögen/Umsatzerlöse)
- Eigenkapitalquote (=Eigenkapital/Gesamtkapital)
- Statischer Verschuldungsgrad  
(=(Verbindlichkeiten+Rückstellungen)/Gesamtkapital)

Nicht berechenbare Kennzahlen (z.B. wegen Division durch 0) werden als 0 angenommen.

Die Ground-Truth kann die Werte 1 oder 0 annehmen, wobei 1 für "Erfolg" und 0 für "Kein Erfolg" steht. Zur Ermittlung der Ground-Truth wurde der Aktienkurs des jeweiligen Unternehmens in den 6 Monaten nach Bilanzstichtag evaluiert und mit dem ursprünglichen Kurs am Bilanzstichtag und dem ATX verglichen:

	Kurs eher > als am Bilanzstichtag	Kurs ca. gleich wie am Bilanzstichtag	Kurs eher < als am Bilanzstichtag
Kurs eher > als ATX	Erfolg	Erfolg	Kein Erfolg
Kurs ca. gleich ATX	Erfolg	Kein Erfolg	Kein Erfolg
Kurs eher < als ATX	Erfolg	Kein Erfolg	Kein Erfolg

#### Besonderheiten:

- Die Vorarlberger Kraftwerke AG (VKW) beschloss im Juni 2013 ein De-listing ihrer Aktie, da nur ein geringer Prozentsatz der Aktien gehandelt wurde. [2]  
Der Datensatz der VKW mit dem Datum 31.12.2012 wurde daher als "Kein Erfolg" angenommen.
- Die Intercell AG fusionierte im Mai 2013 mit Vivalis SA zu dem neuen Unternehmen Valneva SE.[3] Daher wurden die Kursdaten von Valneva SE verwendet.

## 2.4 Feature-Extraktion

Die Feature Extraktion wurde mit der Weka Funktion SVMAttributeEval mit 10-facher Kreuzvalidierung durchgeführt. Die Funktion SVMAttributeEval verwendet einen SVM (Support Vector Machine) Klassifikator, bewertet die einzelnen Attribute mit dem Quadrat ihres SVM-Gewichts und reiht sie, indem sie diesen Wert jedes einzelnen Attributs relativ zu allen anderen evaluiert.[4]  
Folgende Reihung war das Resultat: (*kursive Attribute sind berechnete, normal geschriebene absolute Kennzahlen*)

Ø Rang	+-	Attribut
1	0	<i>Eigenkapitalquote</i>
2,7	0,9	<i>Statischer Verschuldungsgrad</i>
3,3	0,9	<i>Gesamtkapitalrentabilität vor Steuern</i>
3,4	1,11	Betriebserfolg
6,2	4,31	Umlaufvermögen
8,3	1,9	<i>Eigenkapitalrentabilität</i>
8,6	2,84	<i>Deckungsgrad A</i>
8,7	2,05	Anlagevermögen
11,5	2,42	Umsatzerlöse
12,4	5,18	Jahresüberschuss
12,7	2,53	<i>Umlaufintensität</i>
13,3	5,48	EBT
13,4	4,59	Verbindlichkeiten
13,8	2,75	Rückstellungen
15,1	5,24	<i>Kapitalumschlag</i>
15,1	3,56	<i>Sachanlagenbindung</i>
15,2	5,25	Eigenkapital
15,5	2,94	<i>Anlageintensität</i>
17,3	5,57	<i>Vermögensintensität</i>
19,2	3,25	Grundkapital
19,4	3,41	Gesamtkapital
19,6	2,62	Finanzerfolg
20,3	1,68	<i>Umsatzrentabilität</i>

Die Verteilung der absoluten und berechneten Attribute lässt vermuten, dass berechnete Attribute tendenziell aussagekräftiger sind als absolute. Die Mittelwerte der Durchschnittsränge unterstützen diese Vermutung:

Ø-Rang der berechneten Attribute: 10,9

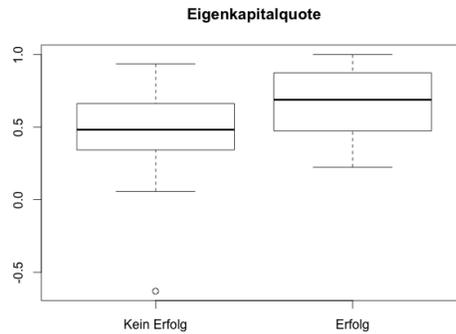
Ø-Rang der absoluten Attribute: 13,01

Es ist möglich, dass bei der Vorhersage von Erfolg an der Börse Relationen von Kennzahlen wichtiger sind als absolute Werte - es sind allerdings weitere, tiefergehendere Analysen notwendig um diese These zu prüfen.

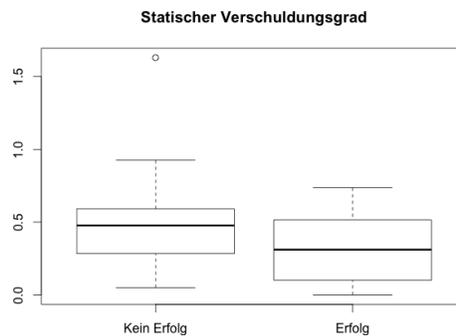
#### 2.4.1 Top 3 Attribute

*Die Boxplots wurden mit RStudio erstellt und zeigen die unterschiedliche Werteverteilung der beiden Klassen "Erfolg" und "Kein Erfolg".*

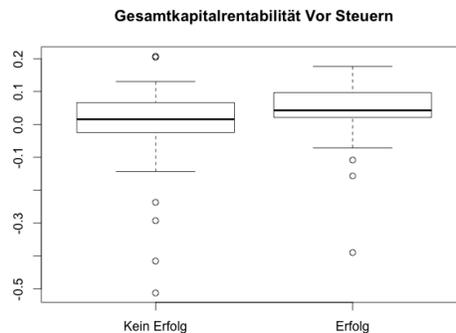
**Eigenkapitalquote** Die Eigenkapitalquote ist der prozentuelle Anteil des Eigenkapitals am Gesamtkapital, und eindeutig die aussagekräftigste Kennzahl zur Vorhersage von Erfolg an der Börse. Viel Eigenkapital bedeutet wenig Fremdkapital und dies ist, zumindest an der österreichischen Börse, wichtig für eine positive Kursentwicklung.



**Statischer Verschuldungsgrad** Der statische Verschuldungsgrad ist der Anteil von Verbindlichkeiten und Rückstellungen am Gesamtkapital, also ein Gegenstück zur Eigenkapitalquote. Ein hoher Verschuldungsgrad korreliert, wie im Boxplot zu sehen ist, mit einer negativen Kursentwicklung.



**Gesamtkapitalrentabilität vor Steuern** Die Gesamtkapitalrentabilität vor Steuern ist das Verhältnis von Betriebserfolg plus Finanzerfolg (=EBT, also Gewinn vor Steuern) zum eingesetzten Gesamtkapital. Auch diese Kennzahl hat eine große Aussagekraft bezüglich der zukünftigen wirtschaftlichen Entwicklung eines Unternehmens.



## 2.5 Erfolg vorhersagen

Zur Vorhersage von Erfolg wurde der SMO Algorithmus von WEKA verwendet. SMO steht für "Sequential minimal optimization" und ist ein Algorithmus der auf dem Support Vector Machine Modell basiert.

Mit dem SMO Algorithmus und **10-facher Kreuzvalidierung**<sup>1</sup> konnten **60%**, mit einer **80/20 Teilung der Daten**<sup>2</sup> sogar **65%** der Testdaten richtig klassifiziert werden.

Wie in nachstehender Tabelle erkennbar ist, werden Testdaten, deren Ground Truth "Erfolg" ist, deutlich öfter richtig klassifiziert als die, deren Ground Truth "Kein Erfolg" ist. Weitere Analysen sind erforderlich um die Gründe hierfür herauszufinden.

	GT: Erfolg	GT: Kein Erfolg	Richtig Klassifiziert
Klassifizierung: Erfolg	39	14	73,58%
Klassifizierung: Kein Erfolg	26	21	44,68%

Tabelle 1: Konfusionsmatrix: Klassifizierung gegen Ground Truth

### 2.5.1 Vergleich: Berechnete gegen absolute Kennzahlen

Das Ergebnis der Feature Extraktion lässt vermuten, dass berechnete Kennzahlen aussagekräftiger sein könnten als absolute. Daher wurde der SMO Algorithmus ein weiteres Mal auf die Testdatensätze angewandt, diesmal allerdings einmal nur mit absoluten und einmal nur mit berechneten Kennzahlen. Die Resultate werden in folgender Tabelle dargestellt:

	absolut	berechnet
10-fache Kreuzvalidierung	53%	59%
80/20 Teilung	40%	50%

Tabelle 2: Richtig vorhergesagte Testdatensätze in %

Auch diese Resultate stützen die These, dass berechnete Kennzahlen generell aussagekräftiger sind als absolute. Im Vergleich zur Vorhersage mit allen Kennzahlen scheidet der Algorithmus hier allerdings deutlich schwächer ab.

### 2.5.2 Top 5 Kennzahlen

Um das Ergebnis der Feature Extraktion weiter zu validieren wurde der SMO Algorithmus nochmals, diesmal nur mit den Top 5 Kennzahlen (siehe 2.4 Feature Extraktion), angewandt.

Sowohl mit 10-facher Kreuzvalidierung, als auch mit einer 80/20 Teilung konnten

<sup>1</sup>Der SMO Algorithmus wird 10 mal durchgeführt, wobei jeweils 1/10 der Daten zum Testen verwendet wird.

<sup>2</sup>80% der Daten werden zum Training, 20% zum Testen verwendet.

65% der Testdaten korrekt vorhergesagt werden. Das ist eine kleine Verbesserung im Vergleich zur Verwendung aller 23 Kennzahlen und unterstützt daher das Ergebnis der Feature Extraktion.

## 2.6 Conclusio

Die Feature Extraktion lässt vermuten, dass berechnete Kennzahlen, für zukünftigen Erfolg an der Börse, tendenziell aussagekräftiger sind, als absolute.

Die drei stärksten Kennzahlen sind Eigenkapitalquote, Statischer Verschuldungsgrad und Gesamtkapitalrentabilität vor Steuern. Jede dieser drei Kennzahlen ist berechnet und zeigt die Relation von jeweils einer anderen absoluten Größe zum Gesamtkapital an.

Mit dem SMO Algorithmus konnten etwa 60% der Testdaten richtig klassifiziert werden. Weitere Versuche mit unterschiedlichen Kennzahlen-Kombinationen unterstützen die Ergebnisse und Erkenntnisse der Feature Extraktion.

## 3 Potenziale für Weiterentwicklungen

- **Mehr Datensätze**

In dieser Arbeit wurden, wie im Abschnitt 2.3 erläutert, 100 Datensätze von 50 Unternehmen verwendet. Die Verwendung weiterer Datensätze würde zu noch genaueren Ergebnissen führen.

- **Bankdaten**

Wie im Abschnitt 2.3 erwähnt, wurden bei dieser Arbeit keine Bankdaten verwendet, da sich Bankbilanzen deutlich von anderen Bilanzen unterscheiden. Das in dieser Arbeit beschriebene Verfahren kann auch auf Bankdaten angewandt werden.

- **Weitergehende Analysen**

Die hier vorgestellten Ergebnisse könnten tiefergehender analysiert werden, um weitere Erkenntnisse zu gewinnen. Beispielsweise könnte untersucht werden, wie groß der tatsächliche qualitative Unterschied in der Aussagekraft von berechneten und absoluten Kennzahlen ist.

- **Andere Börsen evaluieren**

In dieser Arbeit wurden nur Unternehmen der österreichischen Börse berücksichtigt. Sämtliche Ergebnisse sind daher nur für Österreich gültig. Das gleiche Verfahren könnte aber ohne weiteres auf jede beliebige andere Börse angewandt werden. Möglicherweise ist die Aussagekraft von Kennzahlen nicht überall gleich stark.

## Literatur

- [1] <http://issuerinfo.oekb.at>, 18.3.2015
- [2] <https://www.vkw.at/inhalt/at/unternehmen-investoren.htm>, 18.3.2015
- [3] <http://derstandard.at/1363710916654/Der-Name-Intercell-verschwindet-vom-Kurszettel>, 18.3.2015
- [4] <http://weka.sourceforge.net/doc/packages/SVMAttributeEval/weka/attributeSelection/SVMAttributeEval.html>, 22.4.2015