

Einsatz von Data Mining Methoden zur Prognose von studentischem Verhalten

Georg Winterstein

Fachhochschule Mannheim, Fachbereich Informatik
Windeckstraße 110, 68163 Mannheim

g.winterstein@fh-mannheim.de

Zusammenfassung

Kennzahlen spielen für die Verwaltung einer Hochschule sowie für die beschränkt mögliche Auswahl der Studierenden zunehmend eine größere Rolle. Die leistungsbezogene Mittelvergabe berücksichtigt u. a. neben der Gesamtzahl der Studierenden und der Zahl der Absolventen, die Anzahl der Studierenden in der Regelstudienzeit, die Erhöhung des Frauenanteils an der Hochschule sowie die Anzahl ausländischer Studierender in den einzelnen Studiengängen. Bei der Auswahl von Studierenden in Fächern, bei denen eine Zulassungsbeschränkung besteht, ist es für die Hochschule (aber sicher auch für den einzelnen Studienbewerber) interessant, solche Auswahlkriterien zu entwickeln, die eine hohe Studienerfolgswahrscheinlichkeit erwarten lassen. Was liegt also näher, als die an der Hochschule vorhandenen Daten von Studierenden mit Data Mining Methoden zu untersuchen, um sich einen detaillierten Überblick über die momentan bestehende Situation zu beschaffen, sich die Entwicklung der letzten Jahre vor Augen zu führen und eventuell Prognosen zum Studienerfolg bzw. Studienverhalten zu erstellen.

Leider zeigte sich, dass die an der FH Mannheim vorliegenden Daten für das letztere Unterfangen nicht ausreichend und aussagekräftig genug sind, die vorhandenen Daten unzureichend gepflegt und das Interesse der Mitarbeiter sich an der strategischen Planung zu beteiligen auf Grund ihrer Belastung, das operative Geschäft zu bewältigen, gering ist.

Obwohl also zum gegenwärtigen Zeitpunkt das anvisierte Ziel nicht erreicht werden kann, haben aber die Untersuchungen der Daten dazu geführt, dass auf die korrekte und vollständige Datenerfassung mehr Wert gelegt wird, um die Basis für zukünftige Arbeiten zu legen. Des Weiteren konnte durch die Datenanalyse ein genaueres Bild über die momentane Situation gewonnen werden.

1 Datengrundlage

Die Fachhochschule Mannheim verwendet seit 1997 Softwaresysteme der Firma Hochschul-Informationssystem GmbH, Hannover, zur Studenten-, Prüfungs- und Zulassungsverwaltung. Der Modul HISPOS-GX ist das System zur Prüfungsverwaltung, das hochschulspezifische Anforderungen komplexer Studien- und Prüfungsordnungen modelliert und deren Organisation und Durchführung unterstützt. Zu den Anforderungen zählen:

- Abbildung von Prüfungsordnungen
- Zulassungsüberprüfung
- Leistungserfassung
- Notenverbuchung und -berechnung
- Frist- und Terminüberwachung

Mit der universellen Konzeption des Systems geht einher, dass es eine Palette an Anforderungen, Möglichkeiten und Funktionen enthält, die an der Fachhochschule Mannheim nicht bekannt, gebraucht oder verwendet werden. Dazu zählen beispielsweise

- Abwicklung von Freiversuchen
- Prüfergeldabrechnung

Die Datenhaltung erfolgt in der Datenbank SOSPOS, die serverseitig mit INFORMIX, clientseitig mit MS-ACCESS realisiert ist. Die für eine Analyse relevanten studentischen Daten sind im wesentlichen in 3 Tabellen hinterlegt:

SOS Studenten-Stammdaten
STG Studiengangdaten
LAB Prüfungsdaten

Zwischen der Tabelle SOS und STG bzw. LAB bestehen 1:n-Beziehungen. Für einen Studenten gibt es in STG einen Eintrag pro Semester und in LAB einen Eintrag pro Prüfungsleistung, die der Student unternommen hat.

Die Stammdaten werden aus den Tabellen BEW und ANTR gewonnen, die die Daten der Studienbewerber für die einzelnen Studiengänge enthalten. Für diese Tabellen liegen allerdings Einträge erst ab WS 1996 vor. Die früheren Bewerberdaten sind maschinell nicht mehr verfügbar, da sie mit einem proprietären System bzw. manuell bearbeitet wurden. Für die auf Magnetbändern archivierten Daten steht kein Lesegerät mehr zur Verfügung.

Bis auf die Stammdaten werden einmal getätigte Einträge nicht mehr geändert, so dass die Datenbank im wesentlichen wie ein Data Warehouse genutzt werden kann. Die Änderungen in den Stammdaten beziehen sich allerdings nicht nur auf Namens- oder Adressänderungen. Da die Datensätze auch Einträge über Rückmeldung, momentanen Status und Gebührenzahlungen enthalten, wird jeder Datensatz eines Studenten, der sich noch im Studium befindet, im Laufe des Semesters mindestens einmal geändert. Dabei werden nicht nur Felder, die zuvor noch keinen Wert beinhalteten mit Werten gefüllt, sondern es werden auch bestehende Werte geändert.

So ist es z. B. nicht möglich festzustellen, wie oft ein Student im Laufe seines Studiums umgezogen ist oder ob er häufiger die Rückmeldefrist versäumt hat. Die Stammdaten enthalten also für immatrikulierte Studenten deren aktuelle Daten für das laufende Semester, für ehemalige Studenten die des letzten Semesters ihrer Immatrikulation. Die durch diese Vorgehensweise möglicherweise verlorengegangenen Informationen lassen sich auch nicht rekonstruieren, da es für vorangegangene Semester keine Referenzkopien mehr gibt. Außer einer wöchentlichen Datensicherung, die nach ca. 6 Monaten wieder gelöscht wird, werden keine weiteren Kopien der Datenbank gemacht.

Auf Grund eines Programmfehlers wird auch die Information, ob ein Student im laufenden Semester einen Härteantrag gestellt hat, nicht in der Tabelle STG abgespeichert sondern ebenfalls mit den Stammdaten. So kann also auch die Frage, ob ein Student im Laufe seines Studiums ein oder mehrere Härteanträge gestellt hat, nicht beantwortet werden.

Die Tabelle der verfügbaren Studenten-Stammdaten enthält 12453 Einträge. Diese beziehen sich auf Studierende, die ihr Studium im WS 1987 oder später begonnen haben. Diese Daten wurden bei Einführung der HIS-Software aus den vorhandenen Karteikarten 1997 nacherfasst. Wegen Personalknappheit wurde bei der Einführung der HIS-Software entschieden, auf die Erfassung von Studierenden zu verzichten, die vor dem WS 1992 die Hochschule verlassen hatten, auch wenn sie erst nach dem WS 1987 mit dem Studium begonnen hatten.

2 Datenqualität

Die Qualität der vorhandenen Datensätze hat sich seit Erfassungsbeginn bis heute deutlich verbessert. Mögliche Ursachen für fehlerhafte bzw. unvollständige Datenerfassung sind bzw. waren:

- Mit neueren Versionen der Software wurden weitere Attribute eingeführt, deren Werte für bereits existierende Datensätze nicht nachgepflegt wurden bzw. werden können.

- Bei Versionswechsel wurden Defaultwerte offensichtlich gesetzt oder zurückgenommen oder das Design der Eingabemaske wurde geändert.
- In späteren Versionen wurden Auswahlfelder für mögliche Eingaben angeboten, Plausibilitätsprüfungen implementiert oder die Übernahme eines Datensatzes aus der Eingabemaske bei fehlenden Werten verweigert.
- Zum Zeitpunkt, ab dem die Stammdaten der Bewerber dem System zur Weiterbearbeitung zur Verfügung standen, wurden Daten gespeichert, die zuvor nicht erfasst wurden.
- Die Wichtigkeit vollständiger bzw. korrekter Eingaben wurde nicht erkannt bzw. falsch eingeschätzt.
- Aus operativen Gründen wurde der Wertebereich einzelner Felder bei Bedarf im Laufe der Zeit verändert oder erweitert.
- Für die Eingabe nichtnumerischer Werte gibt es keine oder nicht allen bekannte Konventionen.
- Sowohl von Feld zu Feld als auch im zeitlichen Verlauf gibt es Unterschiede in der Behandlung von Nullwerten.
- Eine systematische Überprüfung der Datenbank auf fehlerhafte Eingaben erfolgt nicht.

Darüber hinaus fehlen Informationen, deren Kenntnis für den Studienerfolg für wichtig erachtet werden. So sind z. B. für die Erfassung von Einzelnoten aus der Hochschulzugangsberechtigung (HZB) keine Datenfelder vorgesehen.

3 Auswertungen

Absolventen und Studienabbrecher

Ende des Jahres 2001 waren von den 12453 von der Prüfungsverwaltung erfassten Studierenden 3300 noch immatrikuliert. Das ist der höchste Stand in der Geschichte der Fachhochschule. Die insgesamt vorgefundenen 17 verschiedenen Gründe für eine Exmatrikulation lassen sich zu 3 Gruppen verdichten: Erfolgreicher Studienabschluss, Zwangsexmatrikulation und freiwillige Beendigung des Studiums. Hier ergibt sich folgende Situation

- 3487 Studierende haben einen Abschluss erreicht (38 %)
- 2117 Studierende wurden zwangsexmatrikuliert (23 %)
- 3549 Studierende haben die Hochschule freiwillig verlassen (39 %)

Diese Daten müssen sorgfältig interpretiert werden. Unter den zwangsexmatrikulierten Studierenden sind auch solche, die wegen Überschreitung der Regelstudienzeit das Studium aufgeben mussten. Diese haben jedoch die Möglichkeit, ausstehende Prüfungen (in der Regel die Diplomarbeit) innerhalb eines Jahres abzulegen und somit zu einem Abschluss zu gelangen. Dies konnte für 516 Studierende nachvollzogen werden.

Seit einiger Zeit erhalten Studierende unserer Hochschule bei einem Studiengangwechsel im Haus eine neue Matrikelnummer. Sie erscheinen also in der Tabelle doppelt; für ihren zuerst gewählten Studiengang als Abbrecher, für den zweiten Studiengang als erfolgreiche Absolventen, Abbrecher oder noch im Studium befindlich. Hier finden wir 489 Studierende, die nach einem Studiengangwechsel im neuen Studiengang erfolgreich absolviert haben bzw. die sich in der Nachfrist während ihrer Diplomarbeit in einen anderen Studiengang eingeschrieben haben, ohne zu beabsichtigen, ihn zu beenden, sondern nur, um sich ihren Status als Student zu erhalten. 272 befinden sich nach Abbruch ihres ersten Studiengangs noch im Studium, 646 haben zwei Studiengänge abgebrochen.

Sicher sind auch nicht alle Studierenden, die die Hochschule freiwillig verlassen haben, Studienabbrecher, obwohl ein Hochschulwechsel bei Fachhochschulstudierenden nicht so üblich ist wie für Studierende einer Universität. Insgesamt ist bei 515 Studierenden als Grund für ihren Weggang ein Studienortwechsel angegeben. Eine weitere Sonderrolle spielen 402 Studierende, die im Rahmen eines mit der FH Ludwigshafen durchgeführten gemeinsamen Studiengang Wirtschaftswissenschaften den betriebswirtschaftlichen Teil ihrer Ausbildung in Ludwigshafen absolvieren. In wie weit diese Studierenden ihr Studium erfolgreich absolvieren, entzieht sich unserer Kenntnis. Die bereinigten Zahlen ergeben ein etwas freundlicheres Bild

- 4492 Studierende haben einen Abschluss erreicht (52 %)
- 1601 Studierende wurden zwangsexmatrikuliert (18 %)
- 1684 Studierende haben die Hochschule ohne Angabe von Gründen verlassen (19 %) und
- 917 wegen Wechsels der Hochschule (11 %).

Leider ist nur bei 8 % der exmatrikulierten Studierenden ein Eintrag für die Note der HZB vorhanden, da erst mit Beginn des WS 1997 die Note übernommen wurde. Bis Ende 2001 hat aber noch kein Studierender sein Studium erfolgreich beendet. Der überwiegende Teil der exmatrikulierten Studierenden hat die Hochschule freiwillig nach einem oder zwei Semestern verlassen. Die Frage, in wie weit also aus der Note der HZB auf den Studienerfolg geschlossen werden kann, kann also derzeit noch nicht untersucht werden.

Im Gegensatz zur Note ist bei allen Datensätzen die Art der HZB eingetragen. Von den exmatrikulierten Studierenden besitzen

G. Winterstein

- 52 % die allgemeine Hochschulreife; von ihnen haben 54 % ihr Studium erfolgreich beendet.
- 48 % eine eingeschränkte Hochschulreife; von ihnen haben 49 % ihr Studium erfolgreich beendet.

Wahrscheinlich auf Grund ähnlicher Ergebnisse bevorzugt eine benachbarte Fachhochschule im hochschuleigenen Vergabeverfahren deshalb Bewerber mit allgemeiner Hochschulreife gegenüber Bewerbern mit fachgebundener Hochschulreife.

Die HZB haben

- 6 % der Studierenden im Ausland erworben; von ihnen haben 41 % ihr Studium erfolgreich beendet.
- 94 % der Studierenden im Inland erworben; von ihnen haben 53 % ihr Studium erfolgreich beendet.

Bezüglich der in Deutschland erworbenen HZB gab es die Vermutung, dass der Studienerfolg abhängig ist vom Bundesland, in dem die HZB erlangt wurde. Hier finden wir folgende Situation vor:

Von den exmatrikulierten Studierenden haben ihre HZB erworben:

- 55 % in Baden-Württemberg
- 19 % in Rheinland-Pfalz
- 7 % in Hessen
- 19 % außerhalb dieser Bundesländer

Einen Studienabschluss erzielten

- 60 % der Studierenden mit HZB aus Rheinland-Pfalz
- 49 % der Studierenden mit HZB aus Baden-Württemberg
- 49 % der Studierenden mit HZB aus Hessen
- 52 % der Studierenden, die ihre HZB in einem anderen Bundesland oder im Ausland erworben haben.

In analoger Weise kann man den Studienerfolg auf Grund des Geschlechts der Studierenden bzw. des Ortes, an dem die HZB erworben wurde, untersuchen. Hier ergibt sich folgende Situation:

- 85 % der Studierenden sind männlich. Von ihnen haben 50 % ihr Studium erfolgreich beendet.

- 15 % der Studierenden sind weiblich. Von ihnen haben 60 % ihr Studium erfolgreich beendet.

Von den weiblichen Studierenden, die ihre allgemeine Hochschulreife in Rheinland-Pfalz erworben haben, haben zwei Drittel ihr Studium erfolgreich abgeschlossen.

Eine Auswertung bezüglich der Dauer einer früheren Berufstätigkeit konnte nicht gemacht werden, da diese Daten nur sporadisch eingetragen sind. Dasselbe gilt für die Empfänger von Bafög. Über eine während des Studiums ausgeübte berufliche Tätigkeit liegen ebenfalls keine Daten vor.

Von den Absolventen waren 813 Quereinsteiger, d. h. zu Studienbeginn wurden sie in ein höheres Semester eingestuft. 310 Absolventen haben ein oder mehrere Urlaubssemester im Studienverlauf aufzuweisen. Diese Informationen kann man einem „Reservefeld“ entnehmen, das den Studienverlauf eines Studierenden durch einen String codiert, der die jeweiligen Einstufungssemester enthält.

Der ideale Studienverlauf eines Fachhochschulabsolventen, der einen Diplomstudiengang vollständig an der FH Mannheim absolviert hat, sollte hier durch folgenden Strings abgebildet sein:

- '12P45P78' für Studierende ohne fachspezifische Berufserfahrung
- '1245P78' Studierende, denen das erste praktische Studiensemester erlassen wurde

Der erste Fall trifft für 14, der zweite für 13 Absolventen zu. Die durchschnittliche Studiendauer betrug 9,54 Semester.

Aktuelle Situation

Bei den derzeit immatrikulierten Studierenden ergibt sich folgendes Bild. Von den 3300 Studierenden sind 77 % männlich und 23 % weiblich. 80 % haben die deutsche Staatsangehörigkeit. Allerdings haben 91 % ihre HZB in Deutschland erworben. Die Studierenden mit nichtdeutscher Staatsangehörigkeit kommen aus insgesamt 83 Ländern. Mit mehr als 20 Studierenden sind folgende Länder vertreten:

- Türkei (95)
- Kamerun (76)
- China (46)
- Marokko (44)
- Frankreich (28)
- Bulgarien (22)

Bei den deutschen Studierenden ergibt sich ein stark regionaler Bezug. Von den Studierenden, die eine deutsche Heimatadresse angeben, kommen

- 896 aus Mannheim
- 717 aus dem Rhein-Neckar-Kreis mit Heidelberg
- 321 aus Ludwigshafen Stadt- und Landkreis
- 139 aus dem Kreis Bergstraße
- 130 aus Karlsruhe Stadt- und Landkreis
- 84 aus dem Kreis Bad Dürkheim
- 74 aus dem Neckar-Odenwald-Kreis

Von außerhalb des Verkehrsverbunds Rhein-Neckar kommen 741 Studierende aus insgesamt weiteren 201 Stadt- und Landkreisen Deutschlands.

Im Vergleich finden wir also

Studierende	M / W in %	I / A in %	AH / FH in %
ehemalige	85 / 15	88 / 12	52 / 48
derzeitige	77 / 23	80 / 20	56 / 44

Damit hat die Hochschule sowohl den Anteil von Studentinnen als auch den ausländischer Studierenden steigern können. Hinsichtlich der Verkürzung der Studienzeit lässt sich auch eine leichte Verbesserung erkennen.

WS89	SS90	WS90	SS91	WS91	SS92	WS92
9,6	9,6	9,8	9,8	9,6	9,9	9,7

SS93	WS93	SS94	WS94	SS95	WS95	SS96
9,9	9,7	9,8	9,5	9,3	9,3	9,1

4 Erfolgsprognose

Mit den zuvor einzeln untersuchten Kriterien wurde ein Entscheidungsbaum generiert. Verwendet wurden die binären Eingabevariablen

- Geschlecht (M/W)
- Staatsangehörigkeit (I/A)
- Heimatadresse (I/A)
- Ort der HZB (I/A)
- Art der HZB (AH/FH)
- Erststudium (J/N)
- Urlaubssemester während des Studiums (J/N)

die in fast allen Datensätzen zur Verfügung stehen. Das Ergebnis war nicht sehr aussagekräftig. Als signifikantestes Kriterium stellte sich die Staatsangehörigkeit heraus. Die höchste Wahrscheinlichkeit für ein erfolgreiches Studium lag mit 72 % bei der Gruppe der deutschen, weiblichen Studierenden mit allgemeiner Hochschulreife, die kein Erststudium absolvierten. Die kleine Gruppe der Studierenden ohne deutsche Staatsangehörigkeit, die eine Heimatadresse im Ausland angegeben haben, haben einen Studienerfolg von 66 %. Für die größte Gruppe, Studierende mit deutscher Staatsangehörigkeit im Erststudium ohne Urlaubssemester liegt der Studienerfolg bei 57 %.

5 Zusammenfassung

Das wesentliche Ergebnis der Untersuchung der Studentendaten besteht in der Erkenntnis, dass sich aus den vorhandenen Daten keine entscheidend neuen Erkenntnisse über den Studienerfolg von Studierenden der Fachhochschule Mannheim gewinnen lassen. Eine bessere Datenlage liegt für diejenigen Studierenden vor, die ab 1997 ihr Studium aufgenommen haben. Allerdings haben nur wenige von ihnen ihr Studium bis heute erfolgreich beendet. Weitere, für wesentlich erachteten Daten wie die Einzelnoten aus der HZB, die im Rahmen der Studieneingangsprüfung erhobenen Daten, familiäre Situation, Umfang beruflicher Tätigkeit neben dem Studium etc. können mit dem vorhandenen System nicht erfasst werden und es ist unklar, ob diese überhaupt für alle Studierenden erfragt werden dürfen. Zu untersuchen ist auch, welche, von der individuellen Situation der Studierenden unabhängige Maßnahmen es gibt, die den durchschnittlichen Studienerfolg wesentlich verbessern können.

