

Investigating Success Factors for College Students

Ningning Wu

**Donaghey College of Information Science
and Systems Engineering**

University of Arkansas at Little Rock, AR

Project Goal

The key issue addressed by the project is the investigation of success factors for people in IT related majors

The primary goals

- To develop a sound IT-based methodology that can be applied to any institution of higher education that will help to address the key issue cited above;
- To apply this methodology to the University of Arkansas at Little Rock's Donaghey College of Information Science and Systems Engineering (DCISSE) and obtain fruitful results.

Main Activities

- Developed a new framework (SAZJ) based on commonly used IT tools and applied SAZJ to produce results for this project.
- Identified effective data mining techniques, applied them to our sample database, and found some interesting patterns about the student college performance.

Data Set

**Total 483 Attributes and
8,208 Students' Data**

Age

High School GPA (HSGPA)

SAT or ACT Score

College CGPA

Average Credit Hours Taken (HRTK)

Average Credit Hours Passed (HRPS)

Withdrawn (WD)

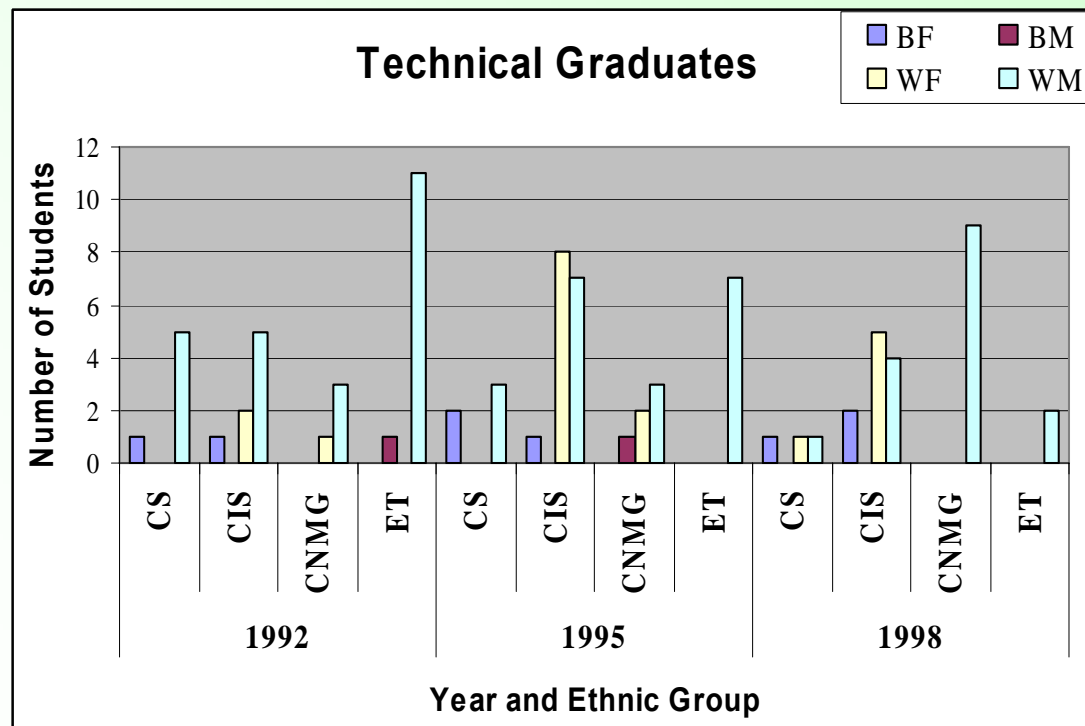
Major

Gender and more

Key Variables 

Findings from SAZJ

- Found a "maturity" factor: People with more experience (as reflected by not having to take the ACT to enter UALR) are more likely to graduate with a technical major.



Findings from Data Mining

- Supervised learning and association rule are two effective techniques
 - Association rule is effective in finding the patterns of highly associated attribute values.

high school gpa > 3 → never withdraw.

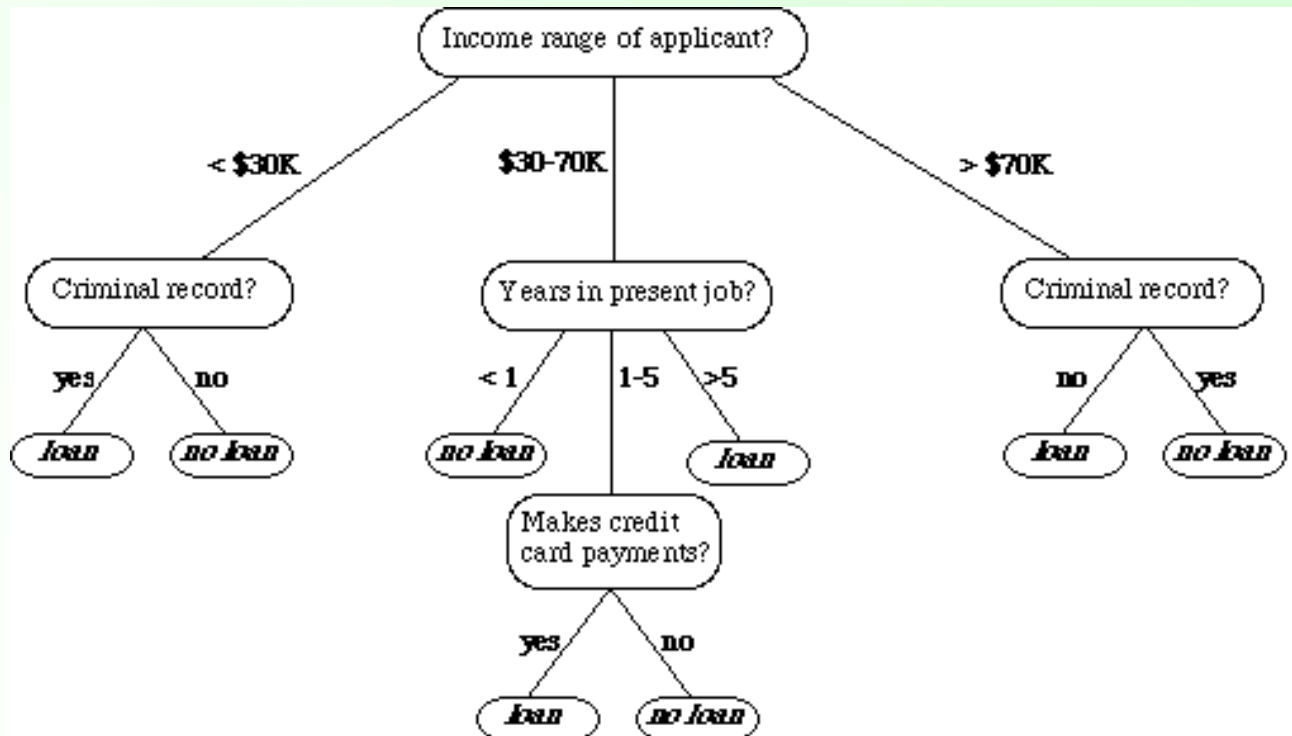
- Decision tree is effective in finding the contrasting and characterizing patterns of different classes.

70% students with less than 3 average attempted credit hours withdrew from a class while less than 8% students with more than 12 average attempted credit hours withdrew from a class.

Decision Tree

It generates a classification model in the form of a tree structure that is either:

- 1) a leaf, indicating a class or
- 2) a decision node that specifies some test to be carried on a single attribute value, with one branch and subtree for each possible outcome of the test.



Association Rules

- Definition:

Given a set of transactions, derives a set of rules in the form of $X \rightarrow Y$, where X and Y are sets of attribute values with

$$X \cap Y = \Phi \text{ and } ||Y|| = 1$$

- Example:

In a market-basket data of supermarket transactions, one may find that customers who buy milk also buy honey in the same transaction, generating the rule

milk \rightarrow honey

- Support

The rule $X \rightarrow Y$ has *support* s in the transaction set T if $s\%$ of transactions in T contain $X \cup Y$

- Confidence

The rule $X \rightarrow Y$ has *confidence* c in the transaction set T if $c\%$ of transactions in T contain X also contain Y

Research Methods

▮ Experiments to measure the effects of Cognitive &

Nocognitive variables:

a) **CGPA in relation to Age, Major, Gender, Hours Taken (HRTK), Hours Passed (HRPS).**

b) **CGPA in relation to SAT, HSGPA, ACT in math, reading and science.**

c) **College Withdrawal in relation to HSGPA, Age, Major, Gender, HRTK, HRPS.**

▮ Student data partitioned into 3 groups: 1) Science & Engineering, 2) Non-Science & Non-engineering, and 3) Undeclared Major.

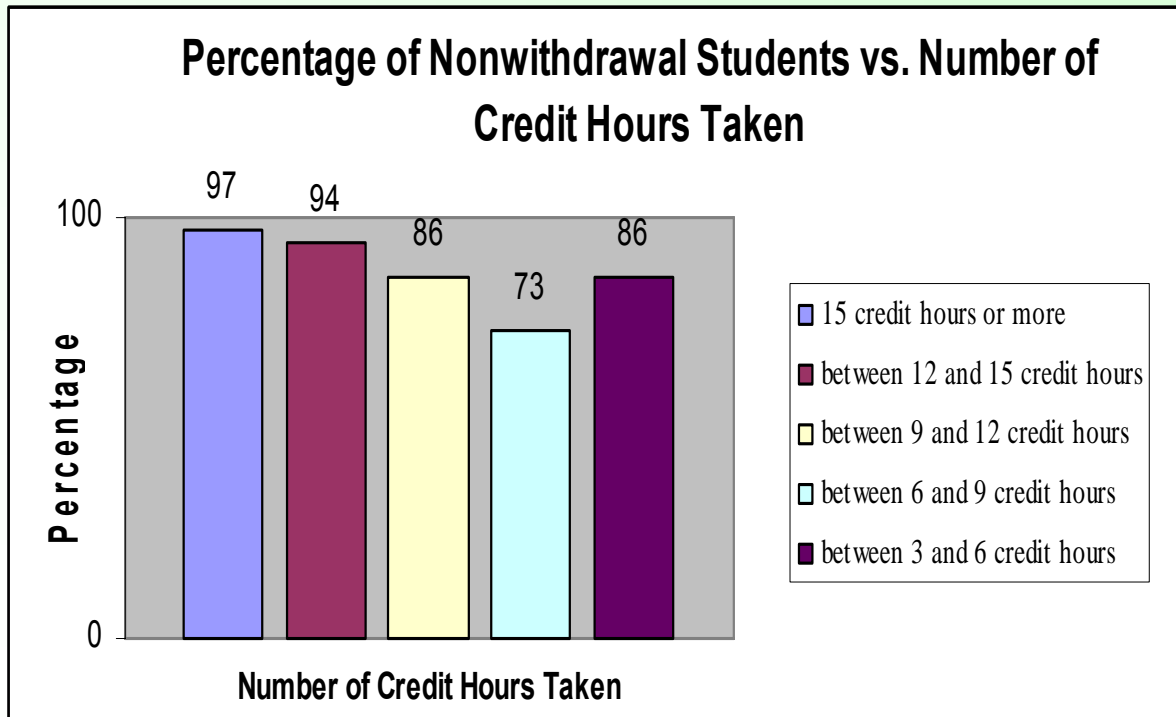
Data Mining Software: WEKA[®] (well recognized, open source, issued under general public license).

Withdrawal Patterns

Science & Engineering Students:

- Students with $3 < \text{high-school-GPA} \leq 4$, $12 < \text{average-cumulative-hours-attempted} \leq 15$, and $20 < \text{ACT-composite-score} < 25$ never withdrew from a class.
- 66% of female students who withdrew at least once were biology major.

Non-Science & Non-engineering Students:



Academic Performance

Science & Engineering Students:

College Activity Based CGPA		
Age	CGPA	Conf.
age \leq 19	cgpa $>$ 3.5	86%
age $>$ 19	cgpa $<$ 1	75%

- 19 yrs or younger \rightarrow CGPA $>$ 3.5; rule's confidence 86%
- 20 yrs or older \rightarrow CGPA $<$ 1; rule's confidence 75%

High School Performance Based CGPA			
Age	HSGPA	CGPA	Conf.
age $<$ 18	*	cgpa $>$ 3.5	86%
18 \leq age \leq 19	1.5 \leq hsgpa \leq 2	3 \leq cgpa \leq 3.5	85%
	hsgpa \geq 2	cgpa \geq 3.5	93%
20 \leq age \leq 49	*	cgpa $<$ 1	75%

- 18 or 19 yrs & 1.5 \leq HSGPA \leq 2 \rightarrow CGPA \leq 3.5; rule's confidence 85%
- 18 or 19 yrs & HSGPA \geq 2 \rightarrow CGPA \geq 3.5; rule's confidence 93%
- Age between 20-49 yrs \rightarrow CGPA $<$ 1; rule's confidence 75%

Academic Performance

Non-Science & Non-engineering Students:

College Activity Based CGPA		
Age	CGPA	Conf.
age \leq 19	cgpa $>$ 3.5	88%
age $>$ 19	cgpa $<$ 1	80%

High School Performance Based CGPA			
Age	HSGPA	CGPA	Conf.
age $<$ 18	*	cgpa $>$ 3.5	88%
18 \leq age \leq 19	1 \leq hsgpa \leq 1.5	2 \leq cgpa \leq 2.5	75%

* Attribute's corresponding value is insignificant or of no influence to outcome

Conclusion

HRTK: the most significant predictor of students' withdrawal trend

Age: the most significant indicator to predict performance of Sc or Eng and Non-Sc or Non-Eng Major students

Future Work

- Apply SAZJ framework and WEKA to the real student database
- Facilitate recruitment and retention efforts

Thanks!