

Анкета участника

Информация	Описание
ФИО студента	Киселёв Евгений Владимирович
Направление/специальность	080800 «Прикладная информатика (в экономике)»
Вуз	Государственное образовательное учреждение высшего профессионального образования Комсомольский-на-Амуре государственный технический университет
Вуз-партнер	Да
Город	Комсомольск-на-Амуре
Кафедра	Информационные системы
ФИО зав. кафедрой	Еськова А.В.
Тема ВКР	АНАЛИЗ ФАКТОРОВ, ВЛИЯЮЩИХ НА ОТЧИСЛЕНИЕ СТУДЕНТОВ, НА ОСНОВЕ ТЕХНОЛОГИИ DATA MINING.
Версия Deductor	5.2
Дата защиты	16.06.2011
Оценка	Отлично
Руководитель ВКР	Канд. т-х. наук, доцент Еськова Анна Владимировна
Представлено	<ul style="list-style-type: none"> ▪ Анкета участника ▪ Аннотация ▪ Пояснительная записка ▪ Данные (приложение пояснительной записке) ▪ Презентация Power Point

Аннотация

АНАЛИЗ ФАКТОРОВ, ВЛИЯЮЩИХ НА ОТЧИСЛЕНИЕ СТУДЕНТОВ, НА ОСНОВЕ ТЕХНОЛОГИИ DATA MINING.

Проблема прерывания обучения в вузе проявляется в разной форме. Во-первых, часть студентов не выдерживает трудностей адаптации к новой среде (в первую очередь речь идет об иногородних студентах). Во-вторых, некоторые студенты разочаровываются в сделанном выборе и меняют место обучения. В-третьих, в семье некоторых студентов изменяются материальные условия, в результате чего студент вынужден либо поменять форму обучения (например, перейти с дневной на вечернюю или заочную) либо прервать обучение, если дальнейшая его оплата становится затруднительной или невозможной.

В данной работе, мы предположим, что процесс прерывания обучения в вузе связан с рядом факторов. К их числу могут относиться проживание в семье или в общежитии, место рождения, платная или бесплатная формы обучения, так же было предположено, что на отсев студентов влияют результаты единого государственного экзамена.

Актуальность исследования заключается в том, что проблема отчисления студентов, является наиболее значимой для администрации ВУЗов, и число отчисленных студентов с каждым годом увеличивается.

Предметной областью исследования являются студенты Факультета Компьютерных Технологий, Комсомольского-на-Амуре Государственного Технического Университета.

Целью выпускной квалификационной работы является сформировать группы студентов, подверженных наибольшему риску отчисления, а так же выявить факторы, влияющие на отчисление студентов, и на их основе предложить возможные варианты повышения эффективности выпуска факультета.

Целью информационной системы является выявление факторов, влияющих на отчисление.

Объектом исследования является конкретный студент, с индивидуальным набором данных, характеризующих его до начала обучения в ВУЗе.

Предметом исследования являются факторы, которые могут повлиять на успешное завершение его обучения.

Основные задачи исследования состоят в следующем:

- определить возможную взаимосвязь вероятностей выбытия из вуза до окончания полной программы обучения и параметров, отражающих неравенство “стартовых позиций” абитуриентов в момент поступления в вуз;
- сформулировать предложения по созданию институциональных инструментов для поддержки уязвимых в экономическом и социальном отношении групп студентов в процессе получения ими высшего образования и снижения риска прерывания этого процесса.

Практическая значимость исследования заключается в том, что с повышением эффективности выпуска студентов, возрастёт престиж факультета, и, следовательно, увеличится количество желающих абитуриентов поступить на него.

Работа состоит из введения, пяти глав, заключения, списка литературы, двух приложений.

Информационно-эмпирическая база данного исследования формировалась на основе данных о студентах, хранящихся в деканате с момента их поступления.

В первой главе, были кратко изложены, основные теоретические сведения об учебно-воспитательном процессе, его основные особенности, цели и этапы его проведения.

Во второй главе были рассмотрены основные характеристики, по которым оценивается успеваемость студентов. Изложены порядки проведения зачётов, экзаменов, а так же текущего контроля в ходе учебного процесса.

В третьей главе представлены сведения об основных технологиях интеллектуального анализа данных. Указаны основные особенности данных и специфика требований к их переработке. Изложены краткие сведения об основных задачах, которые ставятся в ходе обработки больших объёмов данных, таких как классификация и регрессия, позволяющие выявить признаки, характеризующие однотипные группы объектов для того, чтобы по известным значениям этих характеристик можно было отнести новый объект к тому или иному классу. Задача кластеризации, в результате которой происходит группировка объектов, по схожести их свойств, каждый кластер состоит из схожих объектов, а объекты разных кластеров существенно отличаются. Задача прогнозирования, которая на основе текущих и исторических данных оценивает будущие показатели определённых численных данных.

В четвёртой главе рассмотрена постановка задачи, поставленной перед проведением исследования. Был проведён опрос, и выявлены мнения экспертов по вопросу: влияние каких факторов в наибольшей степени затрагивает успеваемость студентов. Но мнения экспертов ничем не подтверждаются, поэтому, для выявления влияния факторов на успеваемость студентов было принято решение воспользоваться аналитической платформой Deductor.

Предположим, что процесс прерывания обучения на факультете связан с рядом факторов. К их числу могут относиться проживание в семье или в общежитии, место рождения, платная или бесплатная формы обучения, так же было предположено, что на отсев студентов влияют результаты единого государственного экзамена.

Для более тщательного анализа причин отчисления студентов, проведём сегментацию студенческой базы. Её целями являются:

- построение кластеров студентов путём выявления их схожей успеваемости;
- оценка наиболее и наименее успешных студентов, и выявление групп подверженных наибольшему риску отчисления.

Эта информация может в дальнейшем использоваться для предугадывания поведения студента, в ходе обучения, и принятия мер, для сохранения максимального числа учащихся, до конца обучения.

Для выявления факторов, наиболее влияющих на отчисление студентов, воспользуемся, корреляционным анализом, который поможет выявить степень линейной функциональной зависимости между входными факторами студентов и их отчислением.

Так же для выявления факторов наиболее влияющих на отчисление студентов воспользуемся моделью множественной логистической регрессией. Коэффициенты в ней оцениваются на основе методов максимального правдоподобия, который позволяет найти такие значения коэффициентов, для которых вероятность появления максимальна.

В пятой главе были описаны этапы проведения исследования. На первом этапе был проведён анализ предметной области. На втором были выделены и описаны основные бизнес-процессы. На третьем этапе была сформирована исходная база студентов, состоящая из 167 человек закончивших обучение в период с 2009 по 2011 годы. На четвёртом этапе были построены самоорганизующиеся карты Кохонена, для разбиения их на кластера.

Вся совокупность была разбита на 5 кластеров.

В первый кластер попали студенты, серьёзно относящиеся к учебному процессу, скорее всего находящихся под контролем родителей, и с ними проживающих. В итоге в первом кластере не наблюдается риска отсева студентов.

Во второй кластер по своим признакам схож с первым, и так же имеет минимальный риск отсева студентов.

Третий кластер является кластером неопределённости. В него попали студенты, находящиеся без контроля родителей, и успех студента в учёбе, зависит только от него. В итоге в третьем кластере присутствует риск отчисления, примерно 50%.

В четвёртый и пятый кластера попали студенты с максимальным риском отчисления.

На пятом этапе, были выявлены факторы, наиболее влияющие на отчисление студентов. С помощью корреляционного анализа, было выявлено, что наиболее значимыми признаками, имеющими линейную функциональную зависимость с выходным полем «Отчисленные», являются: ЕГЭ математика, ЕГЭ русский и место рождения. Таким образом, чем ниже оценки за ЕГЭ, и если студент родился не в городе Комсомольск-на-Амуре, то вероятность отчисления увеличивается.

С помощью модели множественной логистической регрессии были получены следующие отношения шансов:

- «ЕГЭ математика», и её отношения шансов = 0,353, шанс отчисления примерно в 3 раза меньше, чем шанс закончить обучение;
- «ЕГЭ русский», шанс отчисления в 2 раза меньше;
- «ЕГЭ информатика», шанс отчисления примерно в 1,3 раза меньше;
- «ЕГЭ физика», шанс отчисления в 1,5 раза меньше;
- «Место жительства», шанс отчисления в 2,3 раза больше.

Значительно выделяется среди остальных переменная «место рождения». Отношение шансов = 3,096 показывает, что если студент родился в другом населённом пункте, то его шанс быть отчисленным, в 3 раза больше, чем у студента, родившегося в Комсомольске-на-Амуре.

Так же было проведено прогнозирование отчисления студентов, при заданном значении факторов.

На седьмом этапе были предложены методы снижения уровня отчисления.

- Для проблемных групп студентов приравнять консультации к обязательным занятиям, отслеживая посещения;
- организовать курсы выравнивания, для студентов, приехавших из сёл и посёлков;
- ввод кураторства в общежития.

Это далеко не все методы снижения уровня отчисления, но для более обширного анализа, наблюдается большая нехватка исходной информации. Для решения этой проблемы предлагается ввести анкетирование абитуриентов при поступлении в ВУЗ, реквизиты анкеты будут обеспечивать сбор тех данных, которые помогут наиболее широко рассматривать индивидуально каждого студента, и на основе этих данных обеспечивать выявление новых факторов, влияющих на успешное обучение студента в ВУЗ. И если в ходе исследования, влияние данных факторов подтвердится, то факультет сможет легко на них повлиять.

Также был разработан план процесса анализа факторов, влияющих на отчисление студентов.

Процесс анализа факторов, влияющих на отчисление студентов, включает в себя четыре подпроцесса:

1 Осуществление приёмной компании.

В данном подпроцессе происходит приём абитуриентов, и далее получение от них заполненных анкет. После завершения приёмной компании, заполненные анкеты передаются в деканат. Непосредственным исполнителем процесса является представитель приёмной комиссии от факультета.

2 Зачисление студентов, формирование групп.

В подпроцессе происходит обработка анкетных данных, их шифровка и подготовка к виду, который будет использоваться далее при анализе. Исполнителем процесса является секретарь деканата.

3 Анализ групп в течение семестра. Выработка мероприятий.

В данном процессе происходит анализ факторов, влияющих на отчисление студентов, на основе технологии Data Mining. Формируются кластера, и выделяются группы студентов, подверженных риску отчисления. В ходе проведения анализа, вырабатываются мероприятия для сохранения студенческой базы. Исполнителем процесса, является

4 Анализ результатов сессии.

В данном процессе, происходит анализ результатов сессии, и рассмотрение результатов мероприятий, проводимых в ходе семестра. Если мероприятия не дали положительных результатов, происходит их корректировка.