



دانشکده آموزشهای الکترونیکی

پایان نامه جهت اخذ درجه کارشناسی ارشد مهندسی فن آوری اطلاعات

تجارت الکترونیک

کاربرد داده کاوی برای تشخیص گروههای معنادار دانشجویان
دانشگاه آزاد اسلامی به همراه پیاده سازی عملی در یک نمونه واقعی

به کوشش :

صیف اله نیک اندیش

استاد راهنما:

دکتر شهرام جعفری

شهریور ماه ۱۳۹۲

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

به نام خدا

اظهارنامه

اینجانب صیفااله نیک‌اندیش دانشجوی رشته‌ی مهندسی فن آوری اطلاعات گرایش تجارت الکترونیک دانشکده‌ی آموزشهای الکترونیکی اظهارمی‌کنم که این پایان‌نامه حاصل پژوهش خودم بوده و در جاهایی که از منابع دیگران استفاده کرده‌ام، نشانی دقیق و مشخصات کامل آن را نوشته‌ام. همچنین اظهارمی‌کنم که تحقیق و موضوع پایان‌نامه‌ام تکراری نیست و تعهد می‌نمایم که بدون مجوز دانشگاه دستاوردهای آن را منتشر ننموده و یا در اختیار غیر قرار ندهم. کلیه حقوق این اثر مطابق با آیین‌نامه مالکیت فکری و معنوی متعلق به دانشگاه شیراز است.

نام و نام خانوادگی : صیفااله نیک‌اندیش

تاریخ و امضا:
۹۲، ۲، ۲

به نام خدا

کاربرد داده کاوی برای تشخیص گروه‌های معنادار دانشجویان دانشگاه آزاد اسلامی به همراه پیاده سازی
عملی در یک نمونه واقعی

به کوشش

صیف اله نیک اندیش

پایان نامه

ارائه شده به تحصیلات تکمیلی دانشگاه شیراز به عنوان بخشی از فعالیت های تحصیلی لازم برای اخذ درجه
کارشناسی ارشد

در رشته‌ی :

مهندسی فناوری اطلاعات

از دانشگاه شیراز

شیراز

جمهوری اسلامی ایران

ارزیابی کمیته‌ی پایان نامه، با درجه‌ی:

دکتر شهرام جعفری استادیار بخش مهندسی کامپیوتر دانشگاه شیراز (رئیس کمیته)

دکتر مرضیه احمدزاده، استادیار بخش مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی شیراز

دکتر سید مصطفی فخر احمد استادیار بخش مهندسی کامپیوتر دانشگاه شیراز

شهریور ماه ۱۳۹۲

تقدیم به :

پدرم به استواری کوه

مادرم به زلالی چشمه

همسرم به صمیمیت باران

دخترم به طراوت شبانم

سپاسگزاری

سپاس بی‌کران پرودگار یکتا را که هستی‌مان بخشید و به طریق علم و دانش رهنمونمان شد و به همنشینی رهروان علم و دانش مفتخرمان نمود و خوشه چینی از علم و معرفت را روزیمان ساخت. اکنون که این رساله به پایان رسیده است بر خود فرض می‌دانم که از اساتید ارجمند و بزرگووارم

جناب آقای دکتر جعفری که با گشاده رویی و صعه صدر همیشگی رهنمون راهم بودند،

سرکار خانم دکتر احمد زاده که صبورانه هدایت‌گرم بودند،

جناب آقای دکتر فخر احمد که با مهر و اخلاص مثال‌زدنی راهنمایم بودند،

و جناب آقای دکتر مسعودی که زحمات زیادی را متحمل شدند

قدردانی نماییم.

چکیده

کاربرد داده کاوی برای تشخیص گروه‌های معنادار دانشجویان دانشگاه آزاد اسلامی به همراه پیاده سازی عملی در یک نمونه واقعی

به کوشش

صیف اله نیک اندیش

در سال‌های اخیر اکثر واحدهای دانشگاهی دانشگاه آزاد اسلامی با افت دانشجو مواجه شده‌اند. این تحقیق با هدف کنترل نرخ اتلاف و افزایش نرخ موفقیت دانشجویان انجام شد. بدین منظور خروجی‌های موفق آموزشی شامل دانشجویان فارغ‌التحصیل و خروجی‌های ناموفق شامل دانشجویان انصرافی، اخراجی و انتقالی مد نظر قرار گرفت و تلاش گردید با استفاده از تکنیک‌های داده‌کاوی معنادار بودن این گروه‌های دانشجویی تشخیص داده شده و وضعیت آنها برای سالهای آینده پیش‌بینی گردد.

مجموعه داده شامل ۶۰۹۸ نفر از دانشجویانی است که بین سالهای ۱۳۸۳ تا ۱۳۹۱ در دانشگاه آزاد تربت‌جام تحصیل نموده و به شکل موفق یا ناموفق به تحصیل‌شان پایان داده‌اند. متدلوژی مورد استفاده CRISP-DM و تکنیک مورد استفاده رده‌بندی با درخت تصمیم C5 است. روش اعتبار سنجی تقسیم تصادفی داده‌ها با نرم‌افزار داده‌کاوی SPSS Modeler 14.2 به دو مجموعه داده‌ی آموزشی و تست است. و برای ارزیابی قابلیت اطمینان، نتایج این الگوریتم با الگوریتم‌های مشهور دیگر مقایسه شده است.

بر اساس نتایج حاصل، رده‌بندی با درخت تصمیم C5 با دقت ۸۸.۵۷٪ انجام شد و سه گروه دانشجویان انصرافی، اخراجی و فارغ‌التحصیل به عنوان گروه‌های معنادار تشخیص داده شد. و قوانینی برای هر گروه استخراج گردید. همچنین «تعداد کل واحدهای پاس شده» به‌عنوان مهم‌ترین شاخص و «کارآیی علمی در حین تحصیل» به عنوان مهم‌ترین عامل شناخته شد. و اگر چه دقت الگوریتم C5 قابل قبول بود ولی قدرت پیش‌بینی مدل ارائه شده بر اساس CHAID ۹۰/۹۷٪ و نتایج پیش‌بینی مدل ترکیبی SVM و CHAID، ۹۳/۵۱٪ بود.

واژه‌های کلیدی: داده‌کاوی، داده‌کاوی آموزشی، انبار داده، دانشگاه آزاد اسلامی، واحد

دانشگاهی

فهرست مطالب

صفحه

عنوان

فصل اول : مقدمه و طرح مسئله

۲	۱-۱ مقدمه
۴	۲-۱ بیان مساله
۶	۳-۱ اهمیت و ضرورت انجام تحقیق
۷	۴-۱ اهداف تحقیق
۸	۵-۱ کاربردها و استفاده کنندگان از نتایج تحقیق:
۸	۶-۱ ساختار تحقیق
۹	۷-۱ تعریف واژه‌ها و اصطلاحات تخصصی
۱۰	۸-۱ جمع بندی و نتایج فصل

فصل دوم : مبانی نظر تحقیق

۱۲	۱-۲ مقدمه
۱۳	۲-۲ داده کاوی
۱۳	۱-۲-۲ مفهوم داده کاوی
۱۳	۲-۲-۲ تعاریف داده کاوی
۱۴	۳-۲-۲ داده کاوی و کشف دانش از پایگاه داده
۱۴	۴-۲-۲ تاریخچه داده کاوی
۱۵	۵-۲-۲ دلایل پیدایش داده کاوی
۱۵	۶-۲-۲ دلایل نیاز به داده کاوی
۱۷	۷-۲-۲ کاربردهای داده کاوی
۱۷	۸-۲-۲ وظایف داده کاوی
۱۷	۹-۲-۲ مقایسه روش‌های آنالیز آماری و کاوش داده
۱۹	۱۰-۲-۲ کارکردهای داده کاوی
۲۰	۱-۱۰-۲-۲ داده کاوی پیش گوینه
۲۰	۱-۱-۱۰-۲-۲ دسته بندی
۲۰	۲-۱-۱۰-۲-۲ رگرسیون

۲۰	۲-۲-۱۰-۲ داده کاوی توصیفی
۲۱	۲-۲-۱۰-۲ خوشه بندی
۲۱	۲-۲-۱۰-۲ دسته بندی شباهت (قوانین وابستگی)
۲۱	۲-۲-۱۱ چالش های داده کاوی
۲۲	۲-۲-۱۲ محدودیت های داده کاوی
۲۲	۲-۲-۱۳ نرم افزارهای داده کاوی
۲۳	۲-۳ آموزش عالی در ایران
۲۳	۲-۳-۱ تاریخچه
۲۴	۲-۳-۲ انواع موسسات آموزش عالی در ایران
۲۴	۲-۳-۲-۱ بخش دولتی
۲۴	۲-۳-۲-۱-۱ دانشگاه های وابسته به وزارت علوم، و وزارت بهداشت،
۲۴	۲-۳-۲-۱-۲ دانشگاه ها و موسسات وابسته به سایر دستگاه های اجرایی
۲۵	۲-۳-۲-۱-۳ دانشگاه جامع علمی-کاربردی
۲۵	۲-۳-۲-۲ بخش غیردولتی
۲۵	۲-۳-۲-۲-۱ دانشگاه آزاد اسلامی
۲۶	۲-۳-۲-۲-۲ مؤسسات آموزش عالی غیردولتی و غیرانتفاعی
۲۷	۲-۳-۲-۳-۲ بخش غیرحضوری (آموزش از راه دور)
۲۸	۲-۳-۳ آموزش در دانشگاه آزاد اسلامی
۲۸	۲-۳-۳-۱ اخراج از تحصیل
۲۸	۲-۳-۳-۲ انصراف از تحصیل
۲۸	۲-۳-۳-۳ دانشجوی انتقالی
۲۹	۲-۳-۳-۴ فارغ التحصیل
۲۹	۲-۴ جمع بندی و نتایج فصل

فصل سوم: مروری بر تحقیقات انجام شده در رابطه با EDM

۳۱	۳-۱ مقدمه
۳۲	۳-۲ تحقیقات انجام شده در خارج کشور
۳۲	۳-۲-۱ تحقیقات عمومی در رابطه با داده کاوی در آموزش عالی
۳۴	۳-۲-۲ تحقیقات انجام شده در رابطه با ماندگاری دانشجویان
۳۸	۳-۲-۳ تحقیقات انجام شده در رابطه با نرخ گذر دانشجویان
۴۰	۳-۲-۴ تحقیقات انجام شده در مورد نرخ ترک تحصیل دانشجویان
۴۳	۳-۲-۵ تحقیقات انجام شده در رابطه با نرخ موفقیت دانشجویان
۴۴	۳-۲-۶ تحقیقات انجام شده در رابطه با کارآیی علمی دانشجویان

۵۰	۷-۲-۳ تحقیقات انجام شده در سایر زمینه‌های داده‌کاوی آموزشی
۵۴	۳-۳ تحقیقات انجام شده در داخل کشور
۵۴	۱-۳-۳ تحقیقات عمومی در رابطه با داده‌کاوی در آموزش عالی
۵۵	۲-۳-۳ تحقیقات مرتبط با معیار موفقیت دانشجویان
۵۶	۳-۳-۳ تحقیقات انجام شده در زمینه ارزیابی کارایی علمی دانشجویان
۵۸	۴-۳-۳ تحقیقات انجام شده در سایر زمینه‌های مرتبط با سیستم آموزشی
۵۹	۴-۳ خلاصه و جمع‌بندی فصل
	فصل چهارم: روش تحقیق
۶۱	۱-۴ مقدمه
۶۱	۲-۴ فرایند داده‌کاوی
۶۲	۳-۴ مراحل یا گام‌های داده‌کاوی
۶۳	۱-۳-۴ گام درک موقعیت کسب و کار
۶۴	۲-۳-۴ گام شناخت داده‌ها:
۶۴	۳-۳-۴ گام آماده‌سازی داده‌ها
۶۵	۱-۳-۳-۴ پاکسازی داده‌ها
۶۶	۲-۳-۳-۴ یکپارچه‌سازی
۶۷	۳-۳-۳-۴ تجمیع
۶۷	۴-۳-۳-۴ نمونه‌گیری
۶۸	۵-۳-۳-۴ کاهش ابعاد داده
۶۸	۶-۳-۳-۴ انتخاب زیرمجموعه مشخصه‌ها
۷۰	۷-۳-۳-۴ ایجاد مشخصه جدید
۷۰	۸-۳-۳-۴ گسسته‌سازی و باینری‌سازی
۷۱	۹-۳-۳-۴ تبدیل متغیر
۷۱	۴-۳-۴ گام مدل‌سازی
۷۲	۵-۳-۴ گام ارزیابی نتایج
۷۲	۶-۳-۴ گام بکارگیری و توسعه مدل
۷۳	۴-۴ داده‌های تحقیق
۷۴	۱-۴-۴ شیوه و ابزار جمع‌آوری داده‌ها
۷۴	۲-۴-۴ نوع داده‌ها و مقیاس آنها
۷۴	۳-۴-۴ جامعه آماری و نمونه‌گیری
۷۵	۵-۴ اعتبار و کارایی روش ارائه شده
۷۵	۶-۴ ساختار اجرایی تحقیق

۷۶	۱-۶-۴ درک مسئله کسب و کار
۷۶	۲-۶-۴ درک داده ها
۷۷	۳-۶-۴ آماده سازی داده
۷۹	۴-۶-۴ مدل سازی
۸۰	۵-۶-۴ ارزیابی نتایج
۸۰	۶-۶-۴ به کارگیری مدل
۸۱	۷-۴ جمع بندی و نتایج فصل
	فصل پنجم: تجزیه و تحلیل داده ها
۸۳	۱-۵ مقدمه
۸۳	۲-۵ درک مسئله و کسب و کار
۸۳	۱-۲-۵ تعیین اهداف کسب و کار یا تجارت
۸۴	۲-۲-۵ ارزیابی وضعیت
۸۴	۳-۲-۵ تعیین اهداف داده کاوی
۸۴	۴-۲-۵ ایجاد برنامه پروژه
۸۴	۳-۵ درک داده ها
۸۴	۱-۳-۵ انتخاب داده‌های اولیه
۸۶	۲-۳-۵ توصیف داده‌ها
۸۶	۱-۲-۳-۵ نمونه های مورد بررسی پایگاه داده
۸۶	۲-۲-۳-۵ رکوردهای پایگاه داده
۸۷	۳-۲-۳-۵ فیلدهای پایگاه داده
۸۹	۴-۲-۳-۵ نوع و مقیاس داده ها
۹۱	۳-۳-۵ ارزیابی کیفیت داده های تحقیق
۹۴	۴-۵ آماده سازی داده ها برای مدل
۹۴	۱-۴-۵ پاکسازی داده‌ها
۹۵	۲-۴-۵ پالایش داده ها
۹۷	۱-۲-۴-۵ تغییرات داده
۹۷	۲-۲-۴-۵ برچسب گذاری طبقات دانشجویان
۹۸	۵-۵ مدل سازی
۹۸	۱-۵-۵ نحوه فراخوانی داده ها در نرم افزار
۱۰۴	۲-۵-۵ دقت الگوریتم
۱۰۵	۳-۵-۵ اهمیت متغیرها
۱۰۵	۶-۵ اعتبار و کارایی مدل

۱۰۵	۱-۶-۵ ارزیابی اعتبار مدل
۱۰۶	۲-۶-۵ کارایی مدل در مقایسه با الگوریتم های دیگر
۱۱۸	۳-۶-۵ نتایج ترکیب مدل های طبقه بندی
۱۲۱	۷-۵ جمع بندی و نتایج فصل
	فصل ششم: نتیجه گیری
۱۲۳	۱-۶ مقدمه
۱۲۳	۲-۶ نتایج حاصل از تحقیق
۱۲۳	۱-۲-۶ پاسخ به سوالات تحقیق
۱۲۵	۲-۲-۶ نتایج مستقیم حاصل از تحقیق
۱۲۶	۳-۲-۶ مقایسه یافته های پژوهش با پژوهش های قبلی
۱۲۶	۳-۶ پیشنهاد آتی
۱۲۷	۴-۶ محدودیت های تحقیق
۱۲۷	۴-۷ خلاصه و جمع بندی فصل
۱۲۸	فهرست منابع و مآخذ

فهرست جدول‌ها

<u>صفحه</u>	<u>عنوان و شماره</u>
۷۳	جدول ۴ - ۱ وظایف کلی مدل استاندارد CRISP-DM
۸۶	جدول ۵ - ۱ چارچوب کلی پایگاه داده مورد بررسی
۸۷	جدول ۵ - ۲ عناوین فیلدها
۹۰	جدول ۵ - ۳ حالات متغیرهای اسمی
۹۱	جدول ۵ - ۴ مقیاس های عددی
۹۷	جدول ۵ - ۵ یک شکل نمودن داده ها
۱۰۴	جدول ۵ - ۶ درخت تصمیم C5
۱۰۷	جدول ۵ - ۷ شبکه های عصبی Neural Network
۱۰۸	جدول ۵ - ۸ تحلیل تمایزی Discriminant
۱۰۹	جدول ۵ - ۹ رگرسیون لجستیک Logistic
۱۱۰	جدول ۵ - ۱۰ ماشین بردار پشتیبان Support Vector Machine
۱۱۱	جدول ۵ - ۱۱ شبکه های بیزین Bayes Net
۱۱۲	جدول ۵ - ۱۲ درخت تصمیم CART
۱۱۳	جدول ۵ - ۱۳ درخت تصمیم QUEST
۱۱۴	جدول ۵ - ۱۴ درخت تصمیم CHAID
۱۱۵	جدول ۵ - ۱۵ مقایسه تطبیقی صحت الگوریتم ها
۱۱۷	جدول ۵ - ۱۶ مقایسه تطبیقی صحت الگوریتم ها در حالت سه کلاس
۱۲۰	جدول ۵ - ۱۷ نتایج پیش بینی مدل ترکیبی

فهرست شکل‌ها

<u>صفحه</u>	<u>عنوان و شماره</u>
۵	شکل ۱-۱: نمودار گذر یک دانشجو
۶۳	شکل ۴ - ۱ چرخه اجرای پروژه های داده‌کاوی براساس متدولوژی CRISP - DM
۸۸	شکل ۵ - ۱ توزیع داده ها در هر فیلد
۸۹	شکل ۵ - ۲ توزیع طبقات دانشجویان
۹۴	شکل ۵ - ۳ پالایش اولیه و یکپارچه سازی پایگاه داده ها
۹۶	شکل ۵ - ۴ حذف برخی رکوردها
۹۶	شکل ۵ - ۵ حذف برخی فیلدها
۹۸	شکل ۵ - ۶ فراخوانی داده ها در نرم افزار
۹۹	شکل ۵ - ۷ استریم مدل سازی الگوریتم های طبقه بندی
۹۹	شکل ۵ - ۸ استریم درخت تصمیم C5
۱۰۲	شکل ۵ - ۹ درخت تصمیم C5 ایجاد شده
۱۰۵	شکل ۵ - ۱۰ اهمیت شاخص ها
۱۰۷	شکل ۵-۱۱ استریم شبکه های عصبی Neural Network
۱۰۸	شکل ۵-۱۲ تحلیل تمایزی Discriminant
۱۰۹	شکل ۵-۱۳ رگرسیون لجستیک Logistic
۱۱۰	شکل ۵-۱۴ ماشین بردار پشتیبان Support Vector Machine
۱۱۱	شکل ۵-۱۵ شبکه های بی‌زین Bayes Net
۱۱۲	شکل ۵-۱۶ درخت تصمیم CART
۱۱۳	شکل ۵-۱۷ درخت تصمیم QUEST
۱۱۴	شکل ۵-۱۸ درخت تصمیم CHAID
۱۱۶	شکل ۵-۱۹ اولویت بندی الگوریتم ها به لحاظ میزان صحت
۱۱۶	شکل ۵-۲۰ استریم پیش بینی وضعیت دانشجویان جدید
۱۱۹	شکل ۵-۲۱ استریم مدل ترکیبی الگوریتم های طبقه بندی مختلف
۱۲۵	شکل ۶ - ۱ اهمیت شاخص ها

فهرست نشانه‌های اختصاری

EDM	= Educational Data Mining
NCPPE	= National Center for Public Policy and Higher Education
CRISP-DM	= Cross Industry Standard Process for Data Mining
DBMS	= Data Base Management System
DSS	= Decision Support System
MIS	= Management Information System
ACT	= American Collage Test
SAT	= Scholastic Assessment Test
HSG	= High Scholl Grade
SSG	= Senior School Grade
PET	= Performance Evaluation Tool
GIGO	= Garbage In Garbage Out
PCA	= Principal Components Analysis
SVD	= Singular Value Decomposition
FSS	= Feature Subset Selection
KDD	= Knowledge Discovery and Data-mining

فصل اول

مقدمه و طرح مسئله

۱-۱ مقدمه

آموزش یک فاکتور مهم پیشرفت و تعالی یک کشور است، که مردم را به سمت بروز رفتارهای متمدانه سوق می دهد. [۲۲] آموزش عالی نیز یکی از عناصر اصلی جوامع دانش محور محسوب می گردد، امروزه دانشگاهها مهمترین منبع تحقیقات پایه ای و اساسی به حساب می آیند، بنابراین توسعه فناوری های جدید در آنها حیاتی است. [۲۳] در حال حاضر یکی از چالش های مهمی که آموزش عالی با آن مواجه است، رسیدن به مرحله ای است که با داشتن امکانات و تسهیلات لازم به کارآیی و بهره وری بیشتر و فرآیندهای آموزشی دقیق تر برسند. این موسسات در محیطی بسیار رقابتی قرار گرفته اند، لذا درصدد رسیدن به مزیت های رقابتی نسبت به سایر رقبا هستند. و برای باقی ماندن در بازار رقابت به دانش کافی و عمیق برای ارزیابی، برنامه ریزی و تصمیم گیری بهتر نیاز دارند. [۲۴] برای این موسسات که هدفشان مشارکت در بهبود کیفیت آموزش عالی است، موفقیت در زمینه ایجاد سرمایه انسانی موضوعی است که باید بطور مستمر تجزیه و تحلیل گردد، به عنوان نمونه پیش بینی موفقیت دانشجویان برای آنها خیلی مهم است. [۲۵]

دانش مورد نیاز، از نرم افزارهای فعلی مورد استفاده این موسسات قابل حصول نیست، [۲۶] زیرا این نرم افزارها غالباً فقط برای مکانیزه کردن وضع موجود با اجرای پرس وجوهای معمولی و پشتیبانی از مسائل مدیریت و برنامه ریزی کوتاه مدت اداری جوابگو هستند. [۱] از طرف دیگر حجم داده های ذخیره شده در پایگاه داده آموزشی به سرعت در حال افزایش است، این داده ها مربوط به دانشگاه، دانشجویان، اساتید و منابع دیگر است که در اکثر موارد حاوی اطلاعات و الگوهای با ارزشی هستند. [۲۷] داده ها می تواند از سوابق اطلاعاتی و یا داده های عملیاتی موجود در پایگاه داده موسسات آموزشی و یا از سیستم های آموزش مجازی مورد استفاده اکثر موسسات که حاوی اطلاعات زیادی است، جمع آوری گردد. [۲۸-۲۹] همچنین داده ها می تواند علمی و یا مربوط به اطلاعات فردی باشد ولی برای کاربردهایی مثل درک رفتار دانشجویان و کمک به اساتید برای بهبود روش تدریس یا اصلاح دوره های آموزشی و یا سایر موارد بکار رود. [۳۰]

داده کاوی تکنیک استخراج اطلاعات نهان از پایگاه داده بزرگ است. این تکنولوژی قوی با پتانسیل بالا به مراکز آموزش عالی برای تمرکز بر روی اطلاعات مهم موجود در انبارهای داده شان کمک می کند. ابزارهای داده کاوی روندها و رفتارهای آینده را پیش بینی می کند تا به موسسات بطور فعال در تصمیم گیری های دانش محور کمک نماید. البته این تحلیل های خودکار

و آینده‌نگر ارائه شده توسط داده‌کاوی فراتر از تحلیل‌های گذشته‌نگر سیستم‌های پشتیبان تصمیم است. [۳۱] از طرفی دیگر تحقیقات در رابطه با داده‌کاوی در آموزش در حال افزایش است. این زمینه جدید و نوظهور داده‌کاوی آموزشی نامیده می‌شود که با توسعه متدهایی برای کشف دانش از داده‌های مربوط به محیط‌های آموزشی مرتبط است. [۳۲-۳۵] داده‌کاوی به عنوان مناسب‌ترین فناوری به منظور دادن بینش مناسب به اساتید، دانشجویان، دانش‌آموختگان، مدیران و سایر کارکنان، مد نظر است تا این ابزار به عنوان یک دستیار خودکار و بصورت فعالانه به آنها در زمینه تصمیم‌گیری در فعالیت‌های آموزشی کمک نماید. [۳۶-۳۷] همچنین این فناوری می‌تواند به دانشگاه‌ها کمک نماید تا به دقتی بالاتر از ۸۰ درصد (براساس یافته‌ها) پیش‌بینی نماید که کدام دانشجویان موفق به فراغت از تحصیل می‌شوند و کدامیک نمی‌شوند، تا بیشتر بر روی دانشجویان در حال ریسک برای کمک علمی بیشتر تمرکز نمایند. [۳۸]

امروزه در کشورهای در حال توسعه تحقیقات زیادی در زمینه داده‌کاوی آموزشی در حال انجام است، از سال ۲۰۰۴ کنفرانس‌های مهمی از جمله ITS, UM, AIED, ICALT فعالیت خود را در این حیطه گسترش دادند تا اینکه در سال ۲۰۰۸ اولین کنفرانس مستقل داده‌کاوی آموزشی در شهر مونترال کانادا برگزار گردید. [۲] در سالهای اخیر با تحلیل ارتباط بین آموزش عالی و جوامع در مرکز، جنوب و شرق اروپا کارشناسان متعددی در این زمینه در مورد لزوم تجدید نظر در سیستم آموزش عالی مخصوصاً درباره مدیریت استراتژیک موسسات اظهارنظر نموده‌اند. [۳۹]^۱ با افزایش تعداد دانشجو در موسسات آموزش عالی، این موسسات بطور فزاینده‌ای در جهت کارآیی بیشتر حرکت می‌نمایند و اهداف و استراتژی‌هایشان را بر این اساس تنظیم می‌کنند. [۳۹-۴۱] در اروپا جذابیت شاخص‌های کارآیی در حوزه آموزش عالی خیلی زیاد شده است. [۴۲-۴۳] که دلیل اصلی آن تغییرات اجتماعی و سیاسی در سالهای اخیر است؛ دلایلی مثل کاهش تصدی‌گری دولت در نتیجه مشکلات مالی و جذب دانشجو برای آنها، خودمختاری و لزوم شفافیت بیشتر این موسسات، تاثیرگذاری فاکتورهای بیرونی نظیر تغییر نیروی کار در موسسات آموزش عالی در ارتباط بین پژوهش و جامعه. [۴۰، ۴۴-۴۵] همچنین جهانی‌سازی در آموزش عالی باعث افزایش رقابت گردیده است، [۴۵] که باعث تهییج بیشتر دانشجویان و اساتید و جابجایی آنها می‌شود. [۴۶] در ایالات متحده آمریکا نیز مشکلاتی در رابطه با آموزش عالی در سالهای اخیر بروز نموده است، مرکز ملی سیاست‌های عمومی آموزش عالی^۲ (NCPPE) گزارش نمود که نرخ متوسط ماندگاری دانشجویان در ایالات متحده برای سال ۲۰۰۲، ۷۳/۶ درصد بوده است. [۲۸] هر چند این

^۱ - سیزده نفر از کارشناسان در این منبع اظهار نظر نموده‌اند.

^۲ - National Center for Public Policy and Higher Education

مشکل نیز منحصر به ایالات متحده نبوده بلکه بسیاری از کشورها مثل بریتانیا و بلژیک نیز این مشکل را دارند به عنوان نمونه Lau و Vandomme دریافتند که ۶۰ درصد از دانشجویان سال اول و جدیدالورود در بلژیک ترک تحصیل می‌کنند یا اخراج می‌شوند. [۴۸-۴۹]

در کشور ما نیز آموزش عالی با چالش‌های زیادی مواجه است، یکی از چالش‌ها، عدم استفاده از نیروی انسانی، اساتید و متخصصان کشور است. این در حالی است که با توجه به افزایش پذیرش بعضی از دانشگاه‌ها و رشد دانشجویان آنها شاهد افت علمی دانشجویان در هر سال هستیم. [۳] چالش دیگر، افزایش تعداد این موسسات در سالهای اخیر و بدنبال آن کاهش یکباره تعداد دانشجو در بسیاری از این موسسات است، که آنها را دچار مشکل اساسی از نظر کمیت نموده است و مشکل کیفیت آموزشی نیز همچنان یکی از مشکلات حل نشده آنهاست. بنابراین استفاده از تکنیک‌های پیشرفته داده‌کاوی می‌تواند در طبقه‌بندی دانشگاه‌ها، یافتن الگوهای خاص و با ارزش در مورد دانشجویان موفق، یافتن یک برنامه یا روش موفق تدریس، یافتن نقاط بحرانی مالی در مدیریت مالی دانشگاه و موارد دیگر کاربرد داشته باشد. [۲۷]

در این فصل ابتدا مسئله مورد نظر بیان می‌گردد، سپس ضرورت انجام این تحقیق و ناکافی بودن تحقیقات انجام شده در داخل و خارج کشور بررسی می‌شود، همچنین اهداف تحقیق و کاربردها و استفاده کنندگان از آن بیان می‌گردد.

۲-۱ بیان مساله

دانشگاه آزاد اسلامی، دارای فعالیت سی ساله، ۴ میلیون فارغ‌التحصیل، یک میلیون و ۷۰۰ هزار دانشجو، ۴۰۰ واحد دانشگاهی و مرکز آموزشی، [۴] به عنوان یک موسسه آموزشی غیر انتفاعی با نقاط ضعف خاص خود مشغول فعالیت است.

واحد دانشگاهی مورد مطالعه در این تحقیق دانشگاه آزاد اسلامی تربت جام است که در سال ۱۳۶۳ تاسیس شده است و در حال حاضر با ۳۸ رشته شامل ۴ رشته کارشناسی ارشد و ۲۳ رشته کارشناسی و ۱۱ رشته کاردانی و تعداد ۲۸۳۵ دانشجو به عنوان یک موسسه آموزشی در حال فعالیت می‌باشد. [۵]

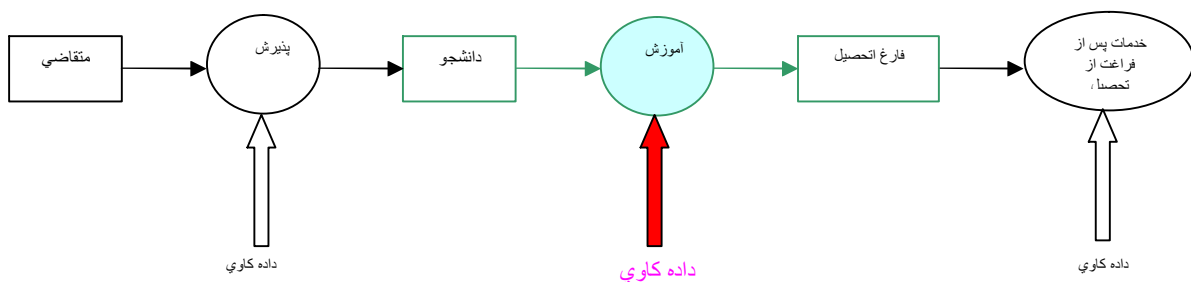
با وجود رشد کمی این موسسه آموزشی، همواره به لحاظ کیفی مشکلات زیادی داشته است. در سالهای اخیر نیز عوامل زیادی ادامه حیات آن را به مخاطره انداخته است بر طبق استراتژی تدوین شده در این موسسه این عوامل شامل نقاط ضعف و تهدیدات محیطی است که شامل فقدان انگیزه جهت انجام امور پژوهشی، توجه بیشتر به مسائل مالی بجای مسائل آموزشی و پژوهشی، بومی نبودن اعضای هیات علمی، عدم وجود مراکز تولیدی و صنعتی متنوع در منطقه، موقعیت جغرافیایی خاص در استان، تنوع مذهبی، عدم وجود زیرساخت‌های ارتباطی و امکانات

رفاهی شهری مناسب، پایین بودن سطح درآمد سرانه منطقه و رشد مراکز آموزش عالی در شهرستان هستند. [۶]

علاوه بر عوامل فوق در سالهای اخیر تعداد دانشجو نیز که به عنوان تنها منبع درآمدی محسوب می گردد در اکثر واحدهای دانشگاهی و از جمله دانشگاه آزاد تربت جام به شدت کاهش یافته و همچنین کیفیت تحصیلی دانشجویان نیز افت نموده است. که می تواند ناشی از تمایل کمتر اقشار جامعه برای تحصیل در این موسسات، عدم شناخت دقیق دانشجویان یا تصویب رشته های نامرتبط بدون توجه به نیاز بازار و یا عدم وجود امکانات مورد نیاز هر رشته و یا جذب هیات علمی نامرتبط با رشته های تصویری و یا سایر عوامل دیگر باشد، [۶] که در نتیجه این مشکلات باعث شده است که این موسسات حتی به روش های غیر معمولی مثل ارسال SMS، حذف آزمون ورودی و ... متوسل شوند.

در دانشگاه آزاد تربت جام با توجه به ثبت نام دانشجویان در ترم های مختلف، حجم انبوهی از داده ها که می تواند حاوی اطلاعات زیادی باشد در پایگاه داده این واحد دانشگاهی ذخیره می گردد. در این تحقیق تلاش می گردد با استفاده از تکنیک های داده کاوی با استفاده از پایگاه داده سیستم مکانیزه دانشگاه آزاد تربت جام به دسته بندی دانشجویان اقدام شود و در صورت امکان وضعیت این گروهها در آینده پیش بینی گردد. تا بتواند در تصمیم گیری بهتر به مدیران این موسسه کمک نماید.

بطور کلی فرآیندهای کلی بکار رفته در سیستم های آموزشی به شرح زیر است.



شکل ۱-۱: نمودار گذر یک دانشجو [یقینی و حیدری، ۱۳۸۷]

کیفیت تخصصی که از این موسسات آموزشی فارغ التحصیل می شوند به کیفیت هر کدام از سه فعالیت (پذیرش و آموزش و خدمات پس از فراغت از تحصیل) بستگی دارد. با تکنیک های داده کاوی و تحلیل داده ها می توان این کیفیت را ارتقاء بخشید. [۵۰]

در این طرح تحقیق، تمرکز بیشتر بر بکارگیری داده کاوی در قسمت آموزش برای تعیین معنادار بودن گروههای مختلف دانشجویان خواهد بود.