Final Report

StudentXpress

Submission by:
Anoop Vintha
Shajay Jayaprakasan
Sukhada Kulkarni
Yashwanth Tekena
INTRODUCTION

The American education system is known for its academic excellence, diversity of education opportunities, cutting-edge technology, flexibility and research opportunities. The education programs in the United States are recognized globally and highly sought after. And hence has become the most preferred and largest destination for international students seeking higher education. Every year thousands of international students travel to the United States in pursuit of academic excellence. Going with the versatility and wide range of programs offered, there are numerous good universities to choose from and often it can get overwhelmingly stressful for a prospective student to choose his/her right university. Our system will ease the life of these students by providing a comprehensive repository of data to help them with their decision making. The system will also have valuable information for the Universities to tap into and gain insights on. Student Xpress will be a one stop destination for students with comprehensive information about universities and their programs. The system will also enable universities to analyze student sentiments and predict trends and tailor program offerings according to student requirements. It will serve as a collaboration point for students and university officials to share and exchange information.

SYSTEM OBJECTIVE

User Needs
- One stop shop for information about universities and courses
- Reliable, accurate and up to date information
- Social, local and mobile accessibility of information

System Functionalities
- Single repository of accurate and up to date information
- Admission probability prediction using “Naïve Bayes” algorithm
- University comparison analytics by student sentiments

Target Customers
- Students
- Universities
- Student Counselors
MARKET LANDSCAPE

There is a lot of data available on the internet. Currently, few websites provide rankings while others have discussion forums. Admission criteria vary each year and can be accessed from university websites. This scattered data makes it difficult to find accurate data at one place.

1. **Major Competitors:**

   - **EducationUSA**
     - https://www.educationusa.info/

   - **Red bus 2 US**
     - http://redbus2us.com/

   - **US News**
     - http://www.usnews.com/rankings

   - **Edulix**
     - http://www.edulix.com/

2. **Gaps in the Market:**

   ✓ No Admission analytics to accurately predict admission.
   ✓ No Geo-based analysis of universities, placements, living expenses, weather, etc.
   ✓ Unstructured distributed data in form of discussion forums.
   ✓ No Course/Course comparison between the universities.

We with our Student Xpress are trying to bridge these gaps. This site will serve as a one-stop-shop for application-related data to better help students who are applying for US educational institutions.
Unique Features

Novelty:

1. Prediction using Naïve Bayes:

We have used Naïve Bayes algorithm to predict whether a student based on his/her profile, would get an admit in a particular university. We used historical student information as the training dataset. This data was crawled from websites, cleaned and processed. The Naïve Bayes algorithm is a probabilistic model based on applying Bayes' theorem and was implemented in the project using Weka.

2. University Comparison Analytics by Student Sentiments:

We have used twitter sentiment analysis using tweets extracted for each university. The tweets are analyzed for their sentiment by passing through SentiWorkNet's lexical dictionary and a score is assigned to each tweet. The scores are then aggregated by University to display the results in graphical charts.
**SYSTEM DESIGN**

The diagram below depicts our systems design. Data from university specific websites are crawled and loaded into the Oracle database which is loaded on university specific web pages. The system also connects to various API’s as shown in the diagram below to retrieve information specific to university the user is interested in. The system is available on desktop, cloud and mobile platforms.

- **Database**: Oracle 11gR2
- **Web Server**: Apache Tomcat
- **Cloud Platform**: EC2 Amazon Cloud Service
<table>
<thead>
<tr>
<th>APIs</th>
<th>Functionalities</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Tableau Logo" /></td>
<td>Tableau is a Analytics tool that allows you to represent data in various graphical visualizations. We used tableau in our project to represent university specific statistical information and results of university specific sentiment analysis</td>
</tr>
<tr>
<td><img src="image" alt="Google Logo" /></td>
<td>We integrated the Google Custom Search API to provide users the ability to search any content on our website</td>
</tr>
<tr>
<td><img src="image" alt="Twitter Logo" /></td>
<td>Twitter API was integrated in our website to pull out tweets specific to universities. This will provide users the ability to see what people are tweeting about a specific university.</td>
</tr>
<tr>
<td><img src="image" alt="Yelp Logo" /></td>
<td>Yelp API was used to pull out information about places, restaurants and houses in an around a specific university</td>
</tr>
<tr>
<td><img src="image" alt="Newsletter Logo" /></td>
<td>Newsletter API allowed our website to send updates and information to customers who have signed up to our website and become members.</td>
</tr>
<tr>
<td><img src="image" alt="Calendar Logo" /></td>
<td>Calendar API allows users of the website to record important events such as application deadlines; exam dates etc and prepare a single consolidated calendar with all events that is of interest to the user.</td>
</tr>
<tr>
<td><img src="image" alt="Google Map Logo" /></td>
<td>Google Map API allows users to get geo based location information specific to universities. Google MAP API is mashed up with the Yelp API to return results with pins.</td>
</tr>
<tr>
<td><img src="image" alt="Zopim Logo" /></td>
<td>Zopim API is customer service/chat API that allows users to communicate with student counselors at StudentXpress.</td>
</tr>
</tbody>
</table>
The Weather API pulls out latest weather information at specific universities that the user looks at.

YouTube API returns videos related universities users access from the StudentXpress website.

Facebook API allows users to log into the system using their Facebook login credentials. Facebook ‘Like’ and ‘Share’ options allow users to share content on social media.

UNIVERSITY SPECIFIC SENTIMENT ANALYSIS PAGE

Below screenshot shows university specific dashboard for sentiment analysis of the tweets. The Dashboard has an overall sentiment pie chart followed by trends over the last one week and positive/negative tweets.
1. University Specific Web Pages:

Following screenshots are web pages specific to universities that students may be interested in.

The below screenshot shows the results of YouTube API pulling out videos specific to the university the user selects. The live chat API that allows users to communicate with student counselors is shown at the bottom right corner of the screenshot.
The screenshot below shows the mash-up of Google map and Yelp API to pull out housing options in the vicinity of the university.

Below screenshot depicts the results of the weather API that shows on the left the current weather in the city and on the right the average temperature in the city over 12 months.
The screenshot below shows the social stream integration with twitter API, Slideshare API and YouTube API. The slideshare API pulls out slides prepared by professors of the specific university a user is looking at.

2. **University Sentiment Comparison Page:**
Below screenshots depict university comparison page by twitter sentiments. The page has two divisions where you can select the university name from a dropdown and the sentiment analysis of that particular university is displayed.
3. Admission Probability Prediction Page:

The admission prediction functionality is a premium feature and requires the user to login and provide his test scores and profile information. The system allows login through face book as well, which can be seen from the screenshot below. Once logged in, the user will be prompt for his test scores and profile information.
Once the necessary information is entered the system returns the probability of admission result by running a prediction model at the back end developed using a training dataset of 4000 records. The model will then predict the probability of admission in that particular university based on the profile information provided.

**Revenue/Business Model**

Our project has a well defined structure of how to create and capture value. Here is our proposed business model to generate revenues from this project:
1. Revenue from universities:

We intend to create login options, where universities can get authorized credentials. Using these credentials universities can get a feedback based on the analysis done on the extracted data. Also students can send anonymous messages which can be passed on to the university authorities. For this access, universities will have to pay annual subscription fees.

2. Revenue from Ads:

It will be beneficial for Ad publishers to post the Ads on our website due to following reason:

- The content will be of high quality, centered around a potential commercial market. The content will be optimized on high-priced keywords.
- Large numbers of students apply each year for education in the US. This will attract a lot of advertisers as the traffic will be high for the site.
Future Directions

5% of applicants = 5000
Per applicant consulting fees = $100
Potential Revenue = $500,000

Plus Revenue from Advertisement Posting

FUTURE DIRECTIONS

- Expand Geographically to other Countries
- Include Online Student Admission Counseling
- Implement on HDFS and Big data technologies
- Expand to include other courses
- Implement More Universities
CONTRIBUTION OF TEAM MEMBERS

- Sukhada Kulkarni
  - Predictive Analytics - Naïve Bayes Algorithm
  - Project Integration
  - Front end Design
  - Business Case

- Anoop Vintha
  - Front End Development
  - Newsletter, Calendar, YouTube and Social Media APIs
  - System Integration

- Yashwanth Takena
  - Backend Development & Integration
  - Mash up of Google Maps & Yelp API
  - Database Administration

- Shajay Jayaprakasan
  - Analytics using Tableau
  - Sentiment Analysis
  - EC2 Web Deployment
  - System Integration

REFERENCES

- https://www.educationusa.info/
- http://www.usnews.com/education
- http://www.edulix.com/
- http://redbus2us.com/