Industry: Sports and Recreation
Company: Dubai Special Olympics
Case Study Link
Fall 2021

Project Goals - Dubai wants to host a special olympics for athletes with various disabilities. They plan to build a detection system that can detect changes in athletics health metrics and diagnose a call to action based on the changes measured. This can ensure a safe event that protects all the athletes. The other main goal of the project is to ensure that both the participants and the viewers will have a good enjoyable time throughout the entire duration of the event.

Analytics Problem Statement - How to predict when a medical emergency is occurring and what are the steps that the medical volunteers can take from there? In addition, how do we ensure that the event is enjoyable for all involved?

There are two processes needed:

1. ***Internet of Things*** - Devices will need to be standardized across the board. We can grant each athlete an arm band for a smartwatch that is capable of the following measurements
    a. Heart Rate
    b. Location
    c. ECG
    d. Heart Rate Variability
    e. Oxygen Saturation
2. ***Data Analytics***

These smart watches must be connected to the same network so all the data can be measured consistently. For these measurements, it will be ongoing refreshes because at any second, the heart rate or other measurement may change after an event or action occurs.

The models that would be used are the following three:

Model 1 - Detect if a particular measurement of heath is higher than a threshold using a CUSUM algorithm

- Note: Since there are different measurements that are being used, we will make individual models. Here is an example, using the metric of heart rate:

Given: Data that will be measured - Heart Rate monitoring of each athlete during hours of games and events

- Data will be collected via Smart Watches.These will be loaned out during registration to standardize the data that is being used

Use: CUSUM method of analysis

To: Detect a decrease or increase in heart rate and will create an alert at a certain threshold either way, reporting a change detected.

Result Interpretation - an alert will be sent to the medical team and next steps will be determined from there. Location on the smartwatch will provide a way to track where the athlete is as well to determine which medical worker should assist based on distance. Additionally, medical workers can determine if the change detected warrants going to the athlete immediately or a check in via Video or Audio call first. If the athlete does not respond to those communication methods, then a medical worker will be released to go meet the athlete.

CUSUM can be set up in such a way that there may be more false detections because the cost of a false detection is far less than the cost of missing a change that is crucial. Since this will be cloud based, there will be a monitoring of steam of data for changes in the incoming observations. Generally speaking, the CUSUM approach will be detecting a shift in the mean of the heart rate measurements. The Log Likelihood CUSUM is a probability testing procedure that sequentially assesses whether the observed outcome is consistent with a specified baseline reference rate.

Model 2 - Prediction of when Health Issues might occur using Logistic Regression Prediction Models.

Given: a various numbers of factors like age of athlete, medical history, gender, what activity/event they would be participating in, weight, and potentially other factors.
Use: Logistic regression on all participants
To: Create a high level distribution of probability to work from to determine the following: number of medical employees needed, number of vehicles and equipment, etc. In addition, we can measure the output of probability by event and determine which locations need more medical workers and which locations would need less.

Model 3 - Clustering to strategically position health personnel, equipment, and vehicles

Given: Data Needed - Previous Special Olympics data to create a classification model off of previous special olympics data; This data will not need to be refreshed frequently as it is static data from the previous year. However, during the actual Olympics if more medical emergencies are happening in one event over another, there may need to be quick adjustments to the setups.
Use: Clustering KNN Modeling **-** The data could be passed through an unsupervised learning algorithm where each issue that occurred would be plotted by date/time that it occurred and by the quantity of events that occurred.
To: From this information, we would be able to start to determine if there are any “clusters” or groupings where medical events tend to occur more often. Based on that information, we can align the time/date data to see what events happened at that time in the previous olympics and determine if certain events are more likely to have medical issues occur and determine site locations for equipment and workers based on that. However, one important note is that the

distribution of risk of last year's data might differ from this year.Hence, we would need to run #2's model on last year's data and this year's data and compare the risk distribution. If they are similar, we can assume the data can be used as is. However, if there is a larger distribution difference, we will need to create an adjustment factor up or down.

Note: Clustering can be used secondarily for determining the number of languages that are necessary on the website for translation. This would be a supervised algorithm, where you can find the clusters with the largest amount of responses and determine how many resources you want to put towards translations. If it is 5 languages for example, you could pick the top 5 clusters.

With my thorough analysis, there was one more component through social media that I felt was necessary for the overall experience of the special olympics.

Model 4 - Social Media Analysis for feedback

Given: all of the data consistently being tracked through the events, use attendee, observer, and participant feedback to make judgement on text data.

- **Data Needed** - Data will not be numeric or in any structured form. It will be text from various social media sources. Hence, the model will be set up to capture certain buzzwords and classify as positive or negative based on that feedback. The immediate dilemma for this approach is that non-numeric data is harder to classify for a machine. In fact, we may have to add an additional layer of deep learning which allows the machines to imitate how humans think and pick up on patterns from the neural inputs.

Use: Sentiment Analysis (opinion mining)
To: determine what % successful the event was overall.

- **Process** - uses NLP and machine learning. Pair social media text data with preset sentiment labels like that of positive, negative, and neutral. The machine can develop underlying sentiments of incoming messages. Image recognition can also be used to determine the reach the event had to capture the number of times the event was discussed via image release. We can measure the total message amount and the percent that was positive and weight the percent that was negative as a deduction to the overall success measurement.

Model 5 - Determining the ads to market the event between various ads
PART A:
Given: different ad options for signing up for the olympics as well as attending the olympics
Use: A/B Testing or a factorial design on ad recommendations
To: Determine the best advertisements and marketing pushes to use to gain more traction for people to watch the events, register to attend, and register to participate

Note: Each of the above, watch the event, register to attend, and register to participate will all have different ads and marketing approaches. Each will have its own factorial design and testing separately. Hence, some may be able to do A/B testing, but some of these marketing

approaches may need to be a large scale factorial design. For instance, if there are a lot of different options/market pushes to get people to register to attend the event, it might be a larger factorial design that needs to be balanced so that each factor is shown the same number of times and tested. A good example is different registration options for different audiences are the following:

1. Register Now! Click Link Here
2. **Register Now! Click Link Here**
3. **Register Now!** Click Link Here
4. REGISTRATION OPEN - Click Link Here!
5. **REGISTRATION OPEN - Clink Link Here!**
6. **REGISTRATION OPEN -** Click Link Here!

*Note - These are just examples of different factors like bolding, wording, capitalization.

PART B
Given: family size, geographic location, hotel choice, age range, etc.
Use: Random Forest Model
To: Produce forest trees that would branch customers into various segments and send out personalized messages depending on the factors given.

Example: If the travelers were a large family, they could send family packages for restaurants and tourist attractions like group biking trails. If the traveler was alone, they could send personalized 1 person activities in the city like a museum tour. Transportation packages can be built based on a combination of factors like group size, location of hotel, age range, etc.

Conclusion

In conclusion, these are just some of the models that were necessary for ***specific*** parts of the entire event. This by no means is all the analytical models that were necessary to put this event on successfully while protecting the health of the participants.