ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

Behavioural Model of Adult Obesity by Childhood Predictors using Crowd Sourcing

Mohd Zareef Affani [#1], Dr. Uma N. Dulhare [#2]

M.Tech Student , Dept. of CSE, Muffakham Jah College of Engineering and Technology, Osmania University, Hyderabad, India[#1].

Professor , Dept. of IT, Muffakham Jah College of Engineering and Technology, Osmania University, Hyderabad, India[#2]

ABSTRACT: Childhood obesity in one of the most serious public health challenges of 21st century. Child development has different stages, so it isn't always easy to know when a child is obese or overweight. Child development refers to the biological, psychological and behavioural changes that occur between birth and end of adolescence. Effective tools are required to determine the behaviours earlier in life and find its influence on weight gain later in life. Crowdsourcing can be used as a tool to assess childhood predictors of adult overweight or adult obesity. In our model, it describe an approach to machine science by allowing the non domain experts to collectively calculate the known and unknown predictors and provide responses to those predictors, such that they are predictive of some behavioural outcome of interest. This was done by building a Web platform and allowing the user to respond to questions which will help to predict a behavioural outcome and it also allow the user to pose new questions. These results in a dynamically building up online survey, but the result of this cooperative behaviour leads to models that can predict the user's outcomes based on their responses to the user-generated survey questions. In our approach we develop a site that will lead to models that can predict user's body mass index. In our approach, it also covers several areas which are identified by earlier research, such as parenting styles, dieting and healthy lifestyle. The results indicate that Crowdsourcing can reproduce already existing hypotheses and also generate new ideas. Users were able to determine the predictors for higher BMI, such as low physical activity in their lifestyle.

KEYWORDS: Crowdsourcing; Machine Science; Behavioural Model; Body Mass Index.

I. INTRODUCTION

There are various cases where one wants to develop a predictive model which can predict the results where a exact solution is not possible. In many cases a model is selected which try to guess the outcome by probability by a given set of input data

A. *Predictive modeling:*

Predictive modelling is a process used in predictive analytics to create a statistical model of future behaviour. It is the area of data mining which is concerned with forecasting probabilities and trends. Predictive modelling can be applied to any type of unknown event, regardless of when it occurred. Regression or neural networks are used to provide methods for finding the model parameters when the predictive variables and the structure of the model have to be taken before. In [2] authors describe how for nonlinear predictive models, the structure of the model is taken from new tools provided from recent works. The task of choosing predictive variables is a qualitative task that requires a domain expertise. For example, in a software engineering, domain knowledge is the knowledge about the environment in which the target system operates. A survey designer must have domain expertise to choose questions that may identify predictive covariates similarly an engineer must develop familiar design in order to get optimize performance.

The domain expert requirement in the system can become a bottleneck to new insights. The knowledge of crowd can be used to produce solutions for difficult problems. Thus, the goal of our research was to test an alternative approach to modelling in which the wisdom of crowds is harnessed to both propose which potentially predictive variables to study by asking questions and to provide the data by responding to those questions. The result is a crowdsourced predictive model.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

B. ***Crowdsourcing:***

Crowdsourcing is a rather anonymous, fast and inexpensive method to generate new hypotheses and discover unexpected issues which might have been overlooked by professionals. Crowdsourcing is the process of getting the work, usually online from a crowd of people. In [7] authors explain that how the growth in user generated information on the Internet is an example of how bottom–up interactions can be used to solve problems that previously required explicit management by teams of experts.

Fig. 1 Crowdsourcing

The process of obtaining the needed ideas services or any content from a large group of people, especially from an online community and harnessing the experience and effort of large numbers of individuals is frequently known as "Crowdsourcing" and has been used effectively in a number of research and commercial applications like in [8] the author explores the possibilities for the model, its potential to exploit a crowd of innovators and its potential for use beyond for profit sectors and the author also explore examples of the model include Threadless, iStockphoto, InnoCentive, the Goldcorp Challenge, and user-generated advertising contests.

For an example of how Crowdsourcing can be useful is, in a primary school children were asked to create a "dream machine" which was then vetted by college design students. Then High school tech students built prototypes for everything from a self made bed to a recycling robot [7].

In our approach, the problem of involvement of domain experts is overcome by introduce a method in which non domain experts that is the general public can be motivated to find the independent variables for predicting the outcome of interest that is body mass index(BMI) for a successful modelling, the goal of this research was to test an alternative approach to modelling in which the wisdom of crowds is harnessed for both purpose, where predictive variables are determined by asking questions and data is collected by responding to those questions. The result is a crowdsourced predictive model.

II. RELATED WORK

In [3] author explore machine science as a new and interesting methodology that uses advanced computational techniques to find, classify and analyse the data, to generate hypotheses and to develop the models. In recent research [4] Robot scientists have demonstrated that can physically carry out experiments [5]. In [2] and [6] the authors explores an algorithms that cycle through hypothesis generation, experimental design, execution, and hypothesis refutation. Except one aspect that has not yet done that is the automation variables selection, for which data should be collected for evaluating hypotheses. In predictive analysis, machine science is unable to select the variables as independent variable that might be used to predict an outcome of interest and for which data collection is required.

Amazon's Mechanical Turk which is a Crowdsourcing tool where, a human describes a "human intelligence task" such as in [9] the authors describe about characterizing data, in [10] the author explains how the Amazon's Mechanical Turk is used for transcribing spoken language and in [11] the authors explains how Crowdsourcing tool, Amazon's Mechanical Turk is used in creating data visualizations. In [12] the authors explain how by involving large groups of humans in many locations, it is possible to complete tasks that are difficult to accomplish with computers alone and would be prohibitively expensive to accomplish through traditional expert-driven processes.

Another well known example for crowdsourced system is the rapid rise of Wikipedia illustrates how online collaboration can be used to solve difficult problems without financial incentives. In [13] Wightman reports that competition is useful in improving performance on a task with either direct or indirect motivation. In our approach, it reports a task with direct motivation that is for the BMI task, users are motivated to understand their lifestyle choices in

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

order to approach a healthy body weight.

In [21]–[23] authors define Citizen Science as the systematic collection and analysis of data, development of technology, testing of natural phenomena and the dissemination of these activities. It is also known as crowdsourced science. In our approach the participants are tried to motivate ideologically. In most citizen science platforms or crowdsourced systems, user contributions are "passive": They contribute computational but not cognitive resources [21], [24]. Some crowdsourced systems allow users to actively participate by searching for items of interest [25] or solve problems through a game interface [26]. In our proposed system, Users are allowed to pose new questions that, when answered by other users, which can be used by a model to predict the outcome of interest.

Understanding the early causes of weight gain has been the focus of a vast amount of research and many determinants of the overweight [14] and obesity have been identified [15]. Finally, this problem can be solved through Crowdsourcing system, which can produce novel creative solutions that will be different from those produced by the domain experts. An iterative crowdsourced poem translation task produced translations that were both surprising and preferable to expert translations [27]. By using the Crowdsourcing system we hope to find the predictive variables which can reveal the unexpected predictors of behavioural outcomes.

III. METHODOLOGY

In our proposed system, it captures the behaviour structure by using the web platform, which has some rules as follows:

- First the investigator will defines behavior based outcome that is to be modeled,
- Second the data are collected from users which are registered to our site,
- Third after answering all the available questions from the site the algorithm generates a model,
- Fourth users are motivated to pose some new questions which may propose new independent variables
- Last the moderator will filter the user generated questions.

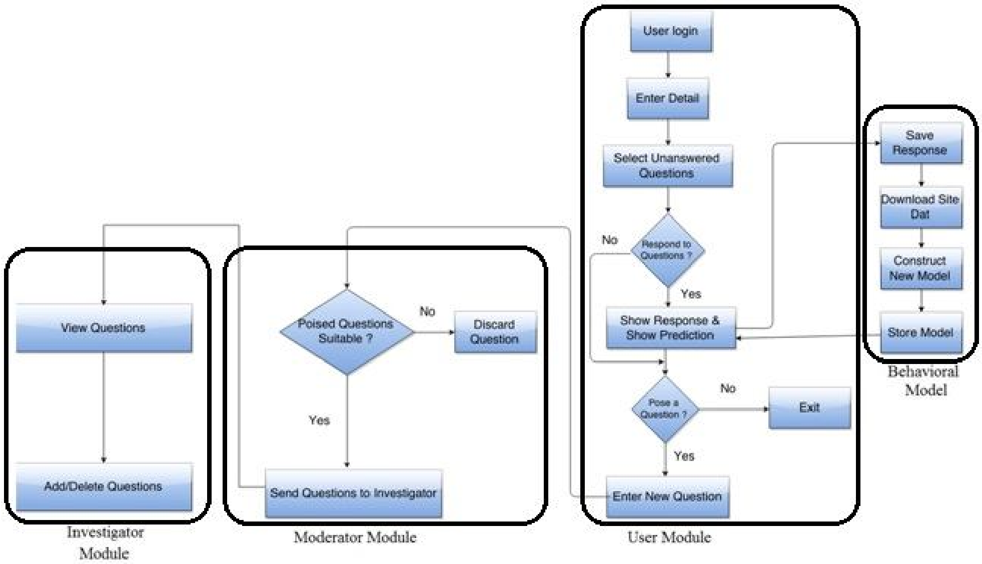

Fig.2. Overview of the Proposed System.

Fig. 2 illustrates how the investigator, moderator, user module and behavioural model work together to produce predictive models of the outcome of interest. The investigator starts the process by constructing a Web site and defines the human based behaviour outcome which has to be modelled then initializes the site by seeding it with a small set of questions which are known to be correlated with the outcome of interest, then the user has to get registered on the site and provide its outcome and start answering the available questions. In our BMI site we start with the question *"When you were a child do you own a bicycle?"* it is a physical activity question, since physical activity plays a major role in healthy lifestyle, for example it may help in reducing the risk of heart diseases, stroke, high blood pressure, type 2 diabetes, osteoporosis and some cancers. The Behavioural model will download the data from the data and constructs a

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

model and store the model parameters.

The participants were allowed to pose their own questions. After posting, these questions are send to the moderator where he will filter those questions by the rules and by checking whether these questions are relevant to our research or not. If this questions are not relevant than this questions will be discarded or if these questions are relevant to our research than this questions will be send to the investigator to add those questions to the site and these questions may be answered by other users which may become new independent variables in the modelling process.

The rules are as follows: A question will be discarded if it is not suitable to our study and if it does not meet:

1. The question revealed the identity of its author (e.g., *"Hi, I am Muneef from India. I would like to know if."*),
2. If the questions contained profanity or hateful text and
3. If the question was inappropriately correlated with the outcome.

Each time participants responds to a question, they were shown a new unanswered question.

In Fig. 3, it explains Crowdsourcing system:

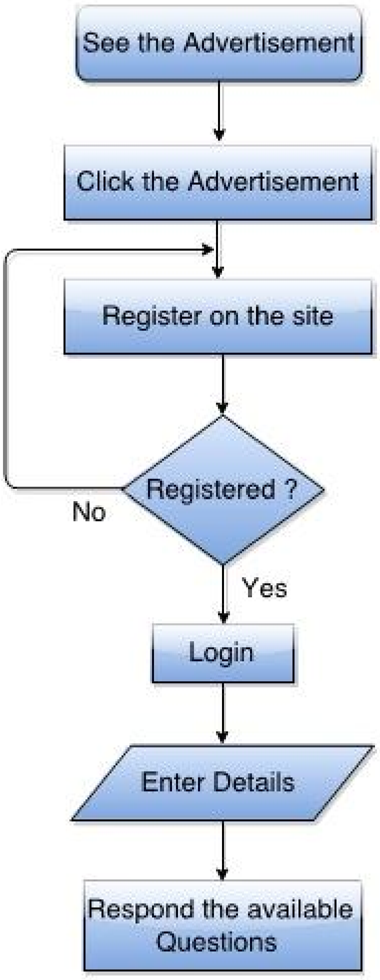

Fig. 3.Process of Crowdsourcing

- First the participants are recruited through the advertisement,
- After clicking the advertisement the participants are taken to our site where,
- They have to first get registered themselves and after getting registered on the site,
- The participants have to login, then they are asked to enter his/her gender, age, height in inches and their weight in kilograms which is later converted to lbs.
- After filling all this information participants Body Mass Index (BMI) were calculated using the eq. (1).
- After the BMI is calculated the participants are directed to answer question page (Fig. 4) where participants has to answer all the available user generated questions including the seeding questions and
- After answering those questions the participants are directed to next page (Fig. 5)where there outcome of interest is shown that is
  - Their actual BMI which is formulated by using the standard formula,
  - The predicted BMI which is calculated by performing regression on all the responses to the questions and
  - A bar graph which shows the comparison between the actual BMI and the predicted BMI in Fig. 5.
  - Depending on the outcome and the BMI classification, the participants are given suggestions about how to achieve ideal weight.

Based on their outcome that may be Underweight, Overweight, Obesity Type I and Obesity Type II.

- Then the participants are allowed to pose their own questions in Fig. 6. With three types of responses:
  - Yes/No,
  - Agree/Disagree,
  - A Number.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

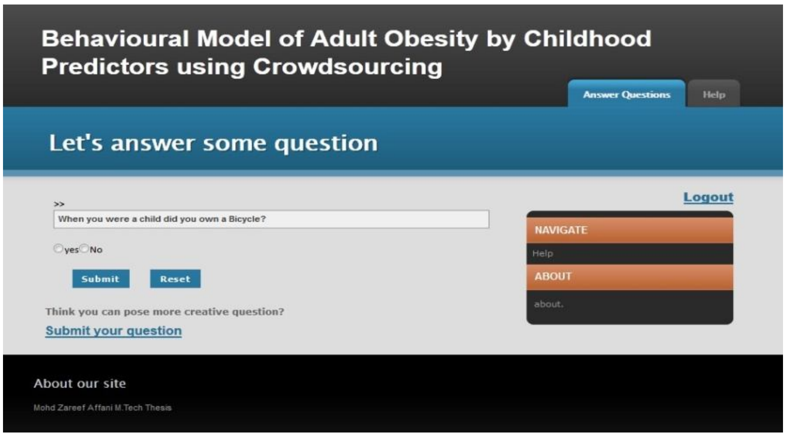

Fig. 4.Answer Question Page

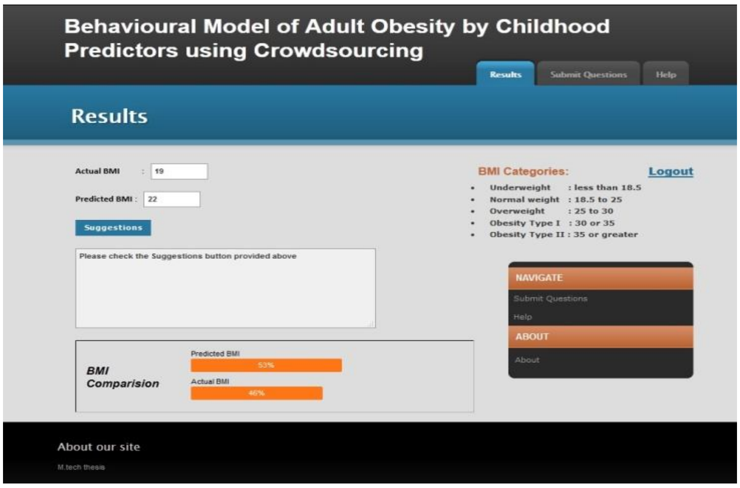

Fig.5. Proposed BMI Result Page

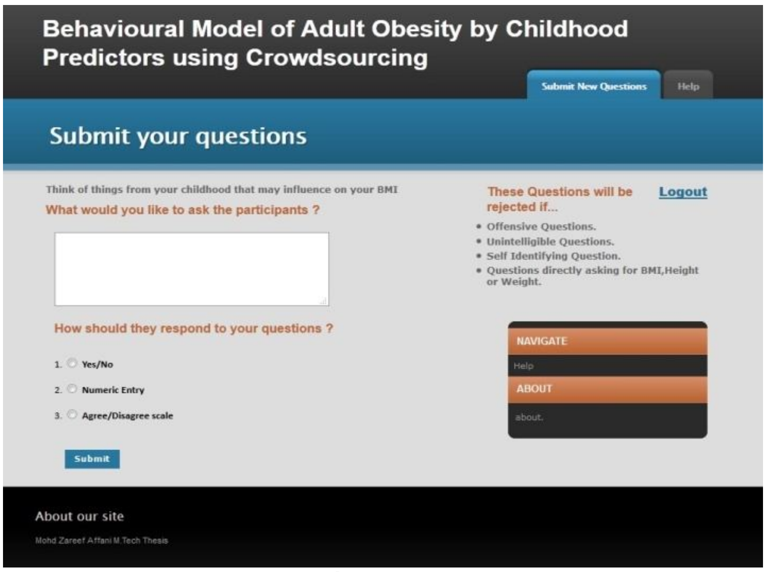

Fig.6. Submit New Question

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

IV. PROPOSED ALGORITHM

A. *Design Considerations:*

- User need to get registered to start the modeling process.
- Enter particular's Height in inches and weight in kilograms (Kg).
- Answer the available questions on the site to get predicted outcome.
- User can pose new questions based on above given rules.

B. *Description of the Proposed Algorithm:*

Aim of the proposed algorithm is to maximize the accuracy of the prediction by using multiple linear regressions and to consider non linear relationship between different variables to model by using the data very efficiently. An additional advantage of our proposed algorithm is that, it is more flexible method where the variables can be numeric or categorical and polynomial terms can also be considered. The proposed algorithm is consists of four main steps.

Step 1: **Calculating General Body Mass Index (BMI):**

Web site was deployed in which models attempted to predict the BMI of each participant, which is still the most common measure for determining a participants level of obesity. Each participant's BMI can be calculated by using *eq. (1)* and this can be done for all the participants.

BMI = **703.3* [Weight * 2.20462] / [height]** ^{2} *eq. (1)*

where weight is in Kilograms and height is in inches.

Step 2: **Calculating Proposed BMI by using Proposed Algorithm**:

In this step, participants have to answer the available questions on the site. After responding to all questions, the values taken in our algorithm are as follows: For categorical variables (yes/no questions) we take 0, 1 value respectively and for ordinal variable like agree/disagree questions we take 0, 1 values respectively and are summed and saved at variable "pbmi" for all these type questions and for numerical questions like "How much sleep did you get daily?" the ideal answer could be 8 hours but if the participants answer is more than the 8 hours than their value is minus from the ideal value and the result value is summed to "pbmi" variable which is then compared to the BMI which is calculated in step 1 and an array is created which contain responses to the questions and the "pbmi" variable and it is taken as input to multiple linear regression which will predict the BMI.

Step 3: **Creating Predictive Models and Calculating Coefficient of determinant:**

In our proposed algorithm, the predictive models are created by using the array which contains participants responses and the "pbmi" variable, by taking this values in two matrixes, the first matrix is $X_{n\times k}$ where n is the number of participants and k is the number of questions and the second matrix is the column matrix Y of length *n* from the collective responses of *n* participants to *k* questions. Each element x_{ij} in X matrix indicates the response of participants *i* to question *j*, and each element y_i in Y indicates the "pbmi" of participants *i*. In the proposed system, multiple linear regressions are used to construct models of the outcome. The modelling process outputs a vector b of length $n + 1$ in *eq. (2)* that contains the coefficients.

$\mathbf{b = (X'X)^{-1}X'Y}$ *eq. (2),* where **X'** is the transpose of matrix **X** and **-1** returns the inverse of the matrix.

It also outputs a vector c of length *k* which is the variance as shown in *eq. (3)* and a vector *d* of length k which is the residual as shown in *eq. (4)*. The coefficient of determinant (r^2) is calculated in *eq. (5)*. These two outputs are then placed in the data store.

$\mathbf{c = (Y'Y) - b'X'Y}$ *eq. (3),* where **Y'** is transpose of matrix Y and **b'** is the transpose of vector b,
$\mathbf{d = Y'Y}$ *eq. (4),* where **Y'** is the transpose of matrix Y which is multiplied with **Y** matrix.
$\mathbf{r^2 = 1 - c/d}$ *eq. (5),* where **c** is the vector which is divided with vector **d** and is subtracted with 1.

Example:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

Q1	Q2	Q3	General BMI	Proposed BMI
4	0	1	27	27
7	1	1	29	27.75
6	1	0	23	24.25
2	0	0	20	18.75
3	0	1	21	22.25

Table 1: General BMI and Proposed BMI values with Q1, Q2 and Q3 values for 5 participants

The body mass index of five people has been measured. For each participant, Q1, Q2 and Q3 values have also been recorded in three columns. The question is: what are the relationships between BMI and Q1, Q2 and Q3? If we have the values of Q1, Q2 and Q3 of a new participant, can we get his/her body mass index? The relationships between BMI and three variables to be something like:

$$\textbf{Proposed BMI} = b_0 + b1 * Q1 + b2 * Q2 + b3 * Q3$$

Based on this equation, in order to predict the value of BMI for a participant with known values of Q1, Q2 and Q3, you need to know the values of b_0, b_1, b_2 and b_3 where b_0 as bias. In most of real-life applications, having a large bias means the predictors (i.e. the three variables) do not have enough predictive power and having small bias is a good sign of having a good predictive model. A large bias could possibly mean that there are descriptors that can explain the observations which we have not discovered them yet.

Here we take BMI column in the table I as column matrix and name it Y and the values of all independent variables as a matrix with name X, by using the eq. (2) we calculate b vector.

```
Matrix X = new Matrix(new double[] []{{4,0,1},{7,1,1},{6,1,0},{2,0,0},{3,0,1}});
Matrix Y = new Matrix(new double[] []{{27},{29},{23},{20},{21}});
MultiLinear ml = new MultiLinear(X,Y);
Matrix b = ml.calculate();
```

Here are the results:

$$b_0 = 9.25,\ b1 = 4.75,\ b_2 = -13.5,\ b_3 = -1.25$$

$$\textbf{BMI} = \mathbf{9.25 + 4.75} * Q1 - 13.5 * Q2 - 1.25 * Q3$$

The size of the values for b and also their sign shows their importance. In this example, Q1 and Q2 have a greater contribution to BMI than Q3, and effect of Q2 and Q1 is opposite, that is participants with high values of Q1 have more BMI. Is this a good model?

As you can see the predicted ones are not that far from the observed ones in table 1. You can find the error (i.e. predicted - observed) for each case and calculate the mean squared error (MSE) which is zero for this example that can indicate how accurate our model is. The lower the MSE, the better the model is. This allows you to calculate r^2 by using equations (3) (4) (5). We get the values of variance that is the vector ‘c’ as “1.5624” and residuals ‘d’ as “2940”. By using these values we get r^2 value as “0.9”.

Step 4: **Showing Suggestions**:

In order to further motivate the participants, in addition to displaying their predicted outcome, participants were also shown suggestions like how to achieve the ideal weight, diet and suggestions about physical activities which should be adopted in their lifestyle. This approach, in effect, provides individualized suggestions to each user as to how slight changes in lifestyle choices and diet may lead to improvements in the health indicator being measured and thus motivate them to formulate new or better questions that might be even more predictive. For example, one user posed the question “*Does your parents were obese when you were a child?*”

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

V. RESULT ANALYSIS

The BMI site was started on Monday, February 9 2015 stayed active for a week at Muffakham Jah College of Engineering And Technology and was deactivated on Sunday, February 15 2015. During that time, it attracted 89 participants who provided their information. In addition to the seed questions supplied by the investigator, participants proposed new questions which were filtered and added to the site. In total, participants provided 1640 responses to the 29 questions. This method effectively provides suggestions to each and every participant like how slight changes in their lifestyle choices may lead to improvements in the health indicator being measured. For example, one participant posed the question "When you were a child did you engage in regular outdoor activity with your family?" another user may realize that the predictive power of this question which achieved an r^2 in the actual experiment of 0.48 and became the eight most predictive question out of a total of 29 may be due to it serving as an indirect measure of the physical activity which is a component of obesity. This may give idea to the participant to pose a new question which may be framed in a better way to eliciting this information, such as "When you were a child, how many hours did you engage in regular outdoor activity with your family?" (This type of question was not actually posed in the experiment).

Table 2 presents a list of questions with highest r^2 value in increasing order. It shows that whether you own a bicycle, number of times a week you visited fast food restaurant and non fast food restaurant, whether someone packed their child a lunch for school and whether the child engaged with their family in regular outdoor activities were strongly related to having a lower BMI later on in life. Family history (e.g., weight of parents and grandparents) and hours of sleep were related to a higher BMI later on in life. The weakest significant predictors appeared to be watching TV while having meals. The questions were from many of the different classes of factors known or known to influence obesity, including parenting (questions 6, 10, 12, 13, 14 and 16), genetic or biological factor (question 3, 15, and 19 [18]), watch TV (questions 18), dietary (questions 2, 4, 7,9, 11 and 20 [17]), and physical activity related (questions 1, 5, 8, 17 [16]).

Participants were recruited from "Muffakham Jah College of Engineering And Technology" and the social networks of the investigators. Fig. 7(a) shows the total number of participants registered on the site for a week and then a rise after until the termination of the experiment. Fig. 7(b) shows the total number of questions and the rapid increase in the number of questions and no significant increase until one participant submits eight new questions on Tuesday, second day of experiment. Fig. 7(c) shows the total number of responses and there was a gradual increase in the number of responses collected on Tuesday. This can be explained by the fact that, although fewer participants visit the site from the second day onward, there are more questions available when they do and thus more responses are supplied by later participants than earlier participants.

There was an increase in responses and questions when few early participants who return to the site and respond to new questions, as shown in Fig. 8. In Fig. 8 the blue pixel indicates that question j was responded by user i and white pixel indicates that the questions were not responded.

	Questions	r^2
1	When you were a child did you own a Bicycle?	0.72
2	When you were a child how many times a week did you eat at a fast food restaurant?	0.67
3	When you were a child were your parents obese?	0.59
4	How many times per week did you eat at a non-fast-food restaurant?	0.57
5	When you were a child were you involved in any competitive sports?	0.55
6	When you were a child Did you often eat late at night?	0.52
7	Apart from breakfast,lunch and dinner,How many times per day did you eat something in between meals?	0.51
8	When you were a child did you engage in regular outdoor activity with your family?	0.48
9	When you were a child Did you drink juice or soda more often than water?	0.41
1	How much sleep did you get daily?	0.37
11	How many times per week did you bring your lunch to school?	0.35
12	Did you drink skim milk much more often than whole milk?	0.34
13	Did you usually eat together with your family?	0.31
14	When you were a child did your parents encourage you to clean your plate?	0.27
15	When you were a child were your grandparents overweight?	0.23
16	Did your parents prohibit you eating certain foods? (e.g.: sweets: sodas: etc)	0.21
17	Did you spend more time playing outdoors than indoors?	0.19
18	How many times a week did you have a meal while watching television?	0.18
19	Was your maternal grandmother obese ?	0.15
20	How many times a week did you eat Sweets?	0.13

Table 2: Questions with increasing order of coefficients of determination (r^2)

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

At the starting of the site for first several hours of the experiment, the behavioral model Fig. 1was runs once an hour. With the increase in site activity, the behavioral model was set to run for every 5 minutes and after some days due to the decrease in the site activity, the behavioral model was set to run once an hour. The Fig. 7(d) show the r^2 value of the multiple regression models as the experiment proceeded during a week. During the starting days of the experiment when there were more participants than questions [see Fig. 7(a) and 7(b)], the models had an r^2 near 0.6. However, before the end of the experiment when there were more participants (89) than questions (29) the models were performing well with an r^2 near 0.9. In Fig. 7(e) show the r^2 value of each question and there is no correlation find between when a question was posed and how predictive it became: The nineteen and the thirteen were the most predictive question.

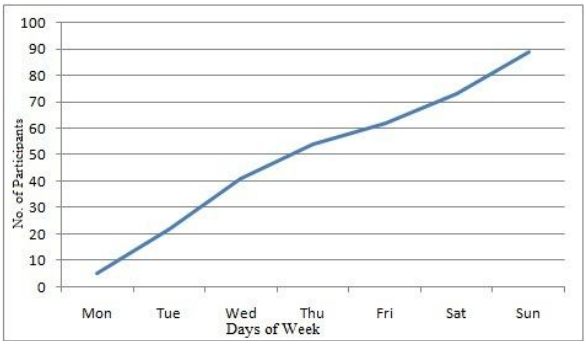

Fig.7(a). Number of Participants

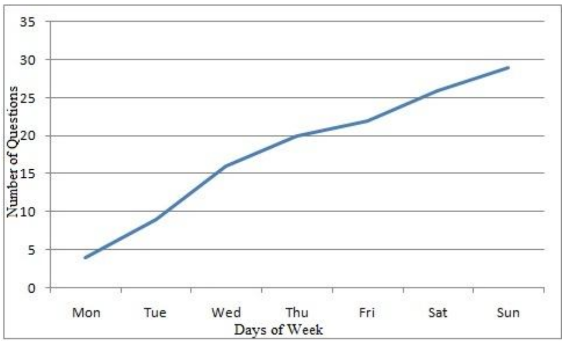

Fig. 7(b). Number of Questions

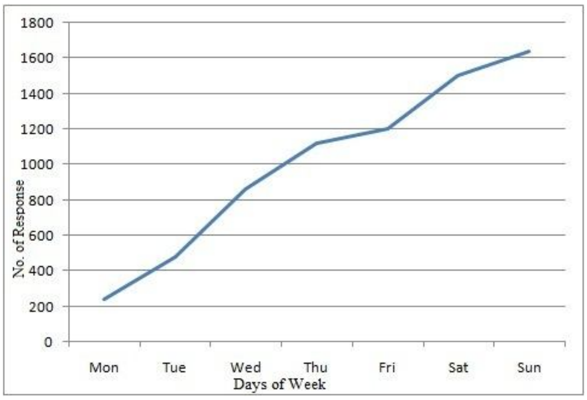

Fig. 7(c). Number of Responses

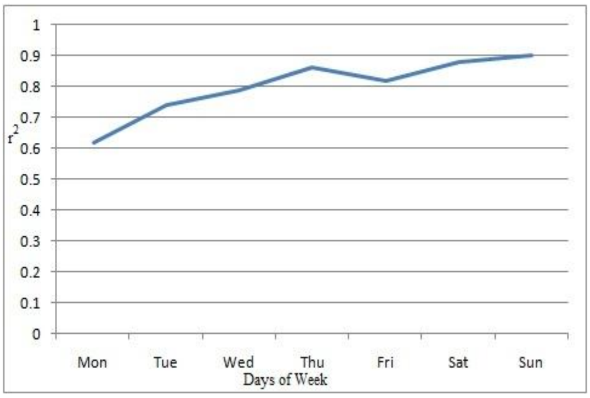

Fig. 7(d). r^2 value day wise

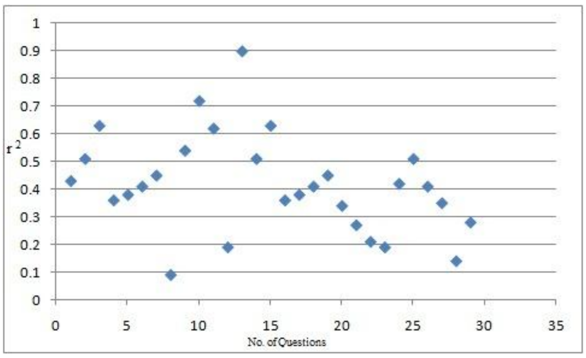

Fig. 7(e). r^2 value of each question

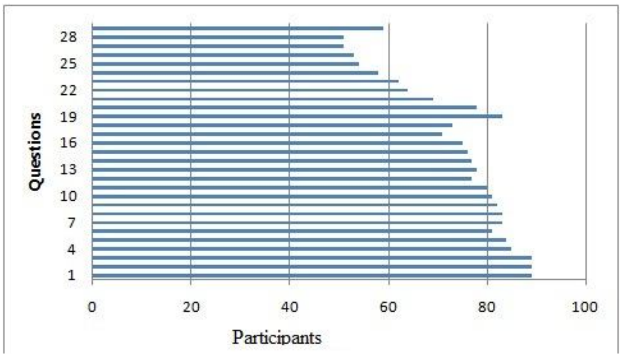

Fig. 8. Participants rate by users of the BMI site

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

The questions posed by the participants are filtered and then those questions were added to site there were only few cases of dishonest responses. Fig. 1 indicates that at least one participant answered the fast food restaurant question dishonestly. The dishonest answer or answers are identified by no individual correlation (r^2 = 0.054) and thus contributes negligibly to the predictions of the models. Questions 1, 2, 3, 6, 7, 8, 10, 15, 19 and 20 as shown in Table I have maximum values and together collected 703 responses, which shows that all responses were not theoretically impossible. This shows that all questions with clear dishonesty (which is defined as giving a response below or above the theoretical minimum or maximum) were quite rare for this experiment.

VI. CONCLUSION

This paper present Crowdsourcing as a efficient tool to determine whether the general public or the crowd could be able to suggest early predictors that are associated with childhood obesity and overweight development. By using the wisdom of the crowd in behavioural research, the Crowdsourcing approach allows non-experts to contribute insight to the research. It is difficult to control the quality of the questions submitted or the demographics of the participants. This work can be only a complement to, rather than a replacement for conventional research methods. The future work could be the insight generated from the Crowdsourcing process can be used to develop new hypotheses, which could be tested in larger more controlled research.

In our proposed system, we overcome the drawback of using linear regression model for capturing all of the explainable variance in the BMI data by using the multiple linear regressions. We conclude that the new predictors discovered in this research were largely related to parenting styles and family environment. Habits learned and initiated in childhood tend to be continued in adult life and therefore a stronger focus should be put on families for establishing healthy habits. Looking at the general family lifestyle may provide broader explanations for the findings of this study. Given that engaging in outdoor activities with family, hours of sleep, and dietary patterns also emerged as significant correlates of BMI, healthy lifestyle during childhood in general is likely to be associated to a lower BMI later on.

REFERENCES

1. Gruber KJ, Haldeman LA "Using the Family to Combat Childhood and Adult Obesity", Preventing Chronic Disease 6: A106, 2009.
2. J. Bongard and H. Lipson, "Automated reverse engineering of nonlinear dynamical systems," Proc. Nat. Acad. Sci., vol. 104, no. 24, pp. 9943–9948, Jun. 2007.
3. J. Evans and A. Rzhetsky, "Machine science," Science, vol. 329, no. 5990, pp. 399–400, Jul. 2010.
4. R. D. King, K. E. Whelan, F. M. Jones, P. G. K. Reiser, C. H. Bryant, S. H. Muggleton, D. B. Kell, and S. G. Oliver, "Functional genomic hypothesis generation and experimentation by a robot scientist," Nature, vol. 427, no. 6971, pp. 247–252, Jan. 2004.
5. R. King, J. Rowland, S. Oliver, M. Young, W. Aubrey, E. Byrne, M. Liakata, M. Markham, P. Pir, L. Soldatova, A. Sparkes, K. E. Whelan, and A. Clare, "The automation of science," Science, vol. 324, no. 5923, pp.85–89, Apr. 2009.
6. J. Bongard, V. Zykov, and H. Lipson, "Resilient machines through continu-ous self-modeling," Science, vol. 314, no. 5802, pp. 1118–1121, Nov. 2006.
7. https://www.wethinq.com/en/blog/2014/08/12/39-Great-Crowdsourcing-Examples.html.
8. D. C. Brabham, "Crowdsourcing as a model for problem solving," Convergence, vol. 14, no. 1, pp. 75–90, Feb. 2008
9. A. Sorokin and D. Forsyth, "Utility data annotation with Amazon Me-chanical Turk," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops, Anchorage, AK, 2008, pp. 1–8.
10. M. Marge, S. Banerjee, and A. Rudnicky, "Using the Amazon Mechanical Turk for transcription of spoken language," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., Dallas, TX, 2010, pp. 5270–5273.
11. N. Kong, J. Heer, and M. Agrawala, "Perceptual guidelines for creating rectangular treemaps," IEEE Trans. Vis. Comput. Graphics, vol. 16, no. 6, pp.990–998, Nov. 2010
12. A. Kittur, E. Chi, and B. Suh, "Crowdsourcing user studies with Mechan-ical Turk," in Proc. 26th Annu. SIGCHI Conf. Human Factors Comput. Syst., Florence, Italy, 2008.
13. D. Wightman, "Crowdsourcing human-based computation," in Proc. 6th Nordic Conf. Human-Comput. Interact.—Extending Boundaries, Reykjavik, Iceland, 2010.
14. Singh AS, Mulder C, Twisk JWR, Van Mechelen W, Chinapaw MJM "Tracking of childhood overweight into adulthood: a systematic review of the literature", Obes Rev 9: 474–488, 2008.
15. Brisbois TD, Farmer AP, McCargar LJ "Early markers of adult obesity: a review", Obes Rev 13: 347–367, 2012.
16. R. Ewing, T. Schmid, R. Killingsworth, A. Zlot, and S. Raudenbush, "Relationship between urban sprawl and physical activity, obesity, and morbidity," Urban Ecology, vol. 18, no. 1, pp. 567–582, 2008.
17. 4 R. Bonow and R. Eckel, "Diet, obesity, and cardiovascular risk," N. Engl. J.Med., vol. 348, no. 21, pp. 2057–2058, May 2003.
18. 4 H A. Herbert, N. Gerry, M. McQueen, I. Heid, A. Pfeufer, T. Illig, Wichmann, T. Meitinger, D. Hunter, F. Hu, G. Colditz, A. Hinney, J Hebebrand, K. Koberwitz, X. Zhu, R. Cooper, K. Ardlie, H. Lyon, N. Hirschhorn, N. M. Laird, M. E. Lenburg, C. Lange, and M F. Christman, "A common genetic variant is associated with adult and childhood obesity," Science, vol. 312, no. 5771, pp. 279–283, Apr. 2006.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

19. P. Boumtje, C. Huang, J. Lee, and B. Lin, “Dietary habits, demographics, and the development of overweight and obesity among children in the Dylan Conger received the B.S. degree in computer science from the Univer- United States,” Food Policy, vol. 30, no. 2, pp. 115–128, Apr. 2005.
20. Y. Wang and M. Beydoun, “The obesity epidemic in the United States— Gender, age, socioeconomic, racial/ethnic, and geographic characteris-tics: A systematic review and meta-regression analysis,” Epidemiol. Rev., vol. 29, no. 1, pp. 6–28, 2007.
21. D. Anderson, J. Cobb, E. Korpela, M. Lebofsky, and D. Werthimer, “SETI@home: An experiment in public-resource computing,” Commun. ACM, vol. 45, no. 11, pp. 56–61, Nov. 2002.
22. J. Cohn, “Citizen science: Can volunteers do real research?” BioScience, vol. 58, no. 3, pp. 192–197, Mar. 2008.
23. J. Silvertown, “A new dawn for citizen science,” Trends Ecology Evol., vol. 24, no. 9, pp. 467–471, Sep. 2009.
24. A. Beberg, D. Ensign, G. Jayachandran, S. Khaliq, and V. Pande, “Fold-ing@home: Lessons from eight years of volunteer distributed comput-ing,” in Proc. IEEE Int. Symp. Parallel Distrib. Process., Rome, Italy, May 2009, pp. 1–8.
25. A. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M. Raddick, R. Nichol, A. Szalay, D. Andreescu, P. Murray, and J.Vandenberg, “Galaxy zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey?” Month. Notices Roy. Astron. Soc., vol. 389, no. 3, pp. 1179–1189, Sep. 2008.
26. S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A.Leaver-Fay, D. Baker, and Z. Popovic,´ “Predicting protein structures with a multiplayer online game,” Nature, vol. 466, no. 7307, pp. 756–760, Aug. 2010.A. Kittur, “Crowdsourcing, collaboration and creativity,” XRDS, vol. 17, no. 2, pp. 22–26, 2010.
27. S. Bowman, S. Gortmaker, C. Ebbeling, M. Pereira, and D. Ludwig, “Effects of fast-food consumption on energy intake and diet quality among children in a national household survey,” Pediatrics, vol. 113, no. 1, pp. 112–118, Jan. 2004.
28. J. Currie, S. DellaVigna, E. Moretti, and V. Pathania, “The effect of fast food restaurants on obesity and weight gain,” Amer. Econ. J., Econ. Policy, vol. 2, no. 3, pp. 32–63, Aug. 2010.
29. A. Romero-Corral, V. Somers, J. Sierra-Johnson, R. Thomas, M. Collazo-Clavell, J. Korinek, T. Allison, J. Batsis, F. Sert-Kuniyoshi, and F. Lopez-Jimenez, “Accuracy of body mass index in diagnosing obesity in the adult general population,” Int. J. Obesity, vol. 32, no. 6, pp. 959–966, Jun. 2008.
30. L. Barness, J. Opitz, and E. Gilbert-Barness, “Obesity: Genetic, molec-ular, and environmental aspects,” Amer. J. Med. Genetics A, vol. 143, no. 24, pp. 3016–3034, Dec. 2007.
31. T. Parsons, C. Power, S. Logan, and C. Summerbell, “Childhood predic-tors of adult obesity: A systematic review,” Int. J. Obesity Rel. Metabolic Disorders, J. Int. Assoc. Study Obesity, vol. 23, pp. S1–S107, Nov. 1999.

BIOGRAPHY

Mohd Zareef Affani is a M. Tech student of Computer Science & Engineering at Muffakham Jah College of Engineering & Technology, Osmania University. He has received Bachelor of Engineering (BE) degree in Computer Science & Engineering from Osmania University, Hyderabad, India in 2013. His research interests are Data Mining, Data warehouse, Cloud Computing and Computer Networks.

Dr. Uma N. Dulhare received the Ph.D. degree in Computer Science & Engineering from Osmania University, Hyderabad in 2013. Currently she is a professor of Information Technology at Muffakham Jah College of Engg. & Tech, Hyderabad. She has received a Best research paper Award in 2010 and received ASDF Global Award for Best Computer Science Faculty of the Year 2013 by the Lt. Governor of Pondicherry. She is advisory board member of Journal of Data Mining & Knowledge Discovery & has published more than 20 peer reviewed papers. Her major research contains Data warehouse and Data Mining, Fault Tolerant System, Neural Networks, Computer forensics and Cloud Computing.