Improvised Deployment of Category Segmentation with Keyword Filtering in News and Social Media using Big Data

Rakesh Babu Sathyanarayanan[1] | Dr.P.Solainayagi[2] | M.Vetripriya[3]

[1]Computer science and engineering, Madha Engineering college,
[2]Assistant professor, Computer science and engineering, Madha Engineering college,
[3]Assistant professor, Computer science and engineering, Madha Engineering college,

Abstract: In this project, news is classified with respect to the category based on the set of keywords assigned to the centralized server to avoid manual segregation of data and to display news to the users. I am using Machine Learning algorithm to principally classify the news according to the category. The keywords like Ball, Bat, Dhoni, Virat Kohli, LBW, etc., and other related words to the Sports Category, actor, hero, theatre, film, etc., related to Cinema category, same way rest of the keywords will be assigned to the corresponding categories. For some words that can be related with 2 or more categories like Pitch which can be related to both Sports as well as music category will be compared based the keywords received in the complete news and will be categorized based on the maximum no of categories received. I have assigned synonyms into this category which is very will be effective in the proper segregation of the messages. I have also added Automatic Alert system in this application about a particular Keyword or its synonyms so that automatic notification either through email or SMS, so that information will be intimated to the customer even he / she misses out to view the News to stay updated in the category he / she has followed.

KEYWORDS: Image Processing, Electronic invoicing, pdftotext, tesseract, tesseract4.

DOI of the Article: https://doi.org/10.46501/IJMTST0707053

Available online at: http://www.ijmtst.com/vol7issue07.html

As per UGC guidelines an electronic bar code is provided to seure your paper

To Cite this Article: Rakesh Babu Sathyanarayanan; Dr.P.Solainayagi and M.Vetripriya[3]. Improvised Deployment of Category Segmentation with Keyword Filtering in News and Social Media using Big Data. *International Journal for Modern Trends in Science and Technology* 2021, *7*, 0707125, pp. 308-318. https://doi.org/10.46501/IJMTST0707053

Article Info. Received: 14 June 2021; Accepted: 12 July 2021; Published: 26 July 2021

I. INTRODUCTION

By using various methods, we can automatically classify news articles for receiving automatic subscription services as well as well as dynamic content filtering and ordering.In this project, we will study the case of news article classification of other contents like education etc., Machinelearning techniques like sentence/text/documentor multi-label classification that are applied is our focus of study. Although many studies are available for English documents or articles, other languages, such as Chinese, still do not have many references.

The wearable devices automate the unpredictable causes of elderly people by treating them or finding the reasons in no time. The smart wearable devices allow the elderly people to get signs or know the causes of various problems like checking heart condition, blood pressure, sugar test, thyroid test, fitness test, etc. These devices will allow getting treatment earlier if the elderly patients need immediate treatment. The results from the devices are accessed by doctors from the hospitals using cloud internet platform services [4]. The real-time sensors will act as the wireless sensor network in sending the patient information to the hospital server. We can also use wearable devices at home in an emergency. Additionally, it will reduce travelling time and cost for monthly or weekly physical check-ups. Hence wearable devices are more important in maintaining the safety of health either inside or outside of home.

We propose a framework that through fine-grained associations with event facets, we can digest reader comments automatically. For example, given the news and comments about the event of Cricket sample of digesting the comments by our framework with respect to event facets and news specificity. The phrases “stump” and “ball” indicate that the identified event facet “Event Facet1” is related to the Cricket which can be found in the content of News. Our framework also determines which event facets, if any, a reader comment contains. For each event facet, the associated news comments are presented for users to understand the specific concerns or opinions of readers about the event facet. For instance, when the user clicks on the button icon “ViewComments” of Event Facet1, a GUI window will pop up and the associated comments are displayed as shown in the bottom left part. The identified comments related to Event Facet1 also talk about the facet of Sport cricket. We call them as event facet-oriented comments. Our system also allows users to visualize the representative event facet phrases to provide a global perception for each event facet. After clicking the button “View Word Cloud”, another GUI window will pop up showing the word cloud as depicted in the bottom right part of Figure 2. The words with larger fonts

Facet1” is related to the Cricket which can be found in the content of News. Our framework also determines which event facets, if any, a reader comment contains. For each event facet, the associated news comments are presented for users to understand the specific concerns or opinions of readers about the event facet. For instance, when the user clicks on the button icon “ViewComments” of Event Facet1, a GUI window will pop up and the associated comments are displayed as shown in the bottom left part. The identified comments related to Event Facet1 also talk about the facet of Sport cricket. We call them as event facet-oriented comments. Our system also allows users to visualize the representative event facet phrases to provide a global perception for each event facet. After clicking the button “View Word Cloud”, another GUI window will pop up showing the word cloud as depicted in the bottom right part of Figure 2. The words with larger fonts dominate the semantic meaning of the corresponding event facet. Such visualization can give users some insights onwhat readers concern

Table 1 Physical Signs Measured by Smart Wearable System

Physical signs	**Sensors**	**Observations**
Electrocardiograph (ECG)	Skin electrodes	Heart rate, heart rate variability
Electroencephalogram (EEG)	Scalp-placed electrodes	Electrical activity of brain, brain potential
Electromyography (EMG)	Skin electrodes	Muscle activity
Blood pressure	Cuff pressure sensor	Status of cardiovascular system, Hypertension
Blood glucose	Glucose meter	Amount of glucose in blood
Galvanic skin response	Woven metal electrodes	Skin electrical conductivity
Respiration	Piezoelectric sensor	Breathing rate, physical activity, inspiration and expiration
Temperature	Temperature probe	Skin temperature, health state

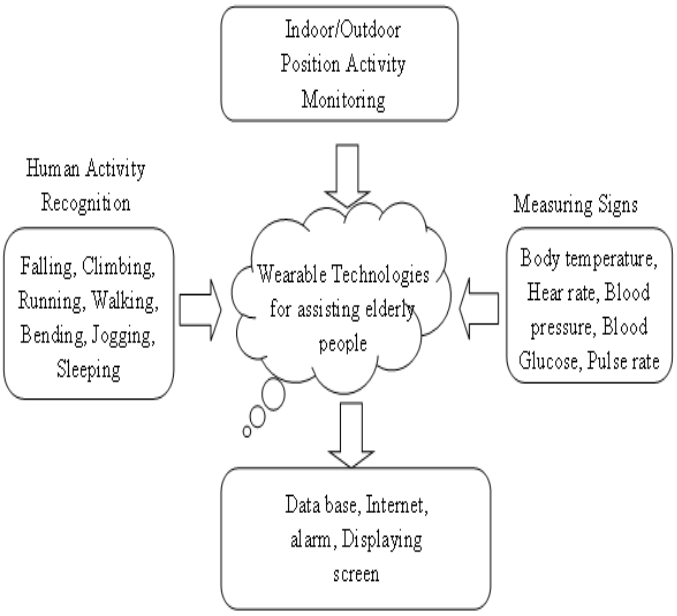


Fig. 1 Wearable Devices in Different Environment

There are many kinds of devices used for measuring the health issues of elderly people with embedded wearable technologies. Some of the devices are smart watch, mobile app, smart phone, online computer, smart dress, smart glass etc. Among these devices smart mobile phones are more attractive to many people. Because, mobile phones can be carried everywhere along with them everywhere they go [8]. Machine learning algorithms are analysed by several researchers to analyse large amount of data and improving the performance of the detection rate in several fields [9] [10]. The general schematic diagram of wearable devices in taking care of elderly system is given in figure 1. The main technological purpose of this system is monitoring indoor physical activity and updating the physical status. The wireless sensor network will be used to track the position of people in real time. Also, the software programming will be used to collect the data, extract the features, design model and make the system recognizable. The integrated sensors will act as the prototype of the network system to find the signs and make decisions quickly.

II. RELATED WORK

This work based on the technologies for the elderly people who are facing stress and complexity to lead their populated without illness. Lensoff - CaravagliaanalysedGerontechnology, dealt with elaborating the situation of the adults surviving in the home, their interaction, flexibility, facing problems and the health condition. They focus on describing the individualistic appeal in the organization in the regular days. It helps to identify the diseases with different medical technologies to overcome the problems during the aging person [11]. In [12], Assistive Technology (AT) is proposed to compare the age group between younger people and elder people with the attitudes of health and flexibility. Elderly people are divided between 72 and 81 age group; it yields to carry all the results for their user's behavior. Once the elderly people find difficulties with their health, immediately they prefer for the AT to get the benefits from the sickness and deformity from frightening of the problems.

In [13], the communication heritage condition reaches the new authorization and was developed to measure, provocation and give out the "native" tradition. This mediated communication technology is proposed to distribute the structure of elements that assist the consideration of the communal system, new design, and spectators. It supports the AT for all the communication in the environment for adult's health and follows the interactive adoption model. In [14], demonstrates the usage of the email to communicate with the 18 groups to resolve the innovation and prompts by the old age people. They innovate how it is possible to learn about new ideas for the old era. Is it acceptable by the adults to adopt the users by the communication methods? Hence the internet demands the cost for all usage of the technology that was adopted by older people.

This paper contributes to the Technology Acceptance Model (TAM) to examine the adoption of the internet by the Chinese older adults and how they are involved in their different features. Proposed method is used for the resolution to estimate the senior in younger and elder involvement for comparing different factors with the TAM model. It consists of many features and factors such as facilitating conditions (FC), perceived ease of use (PEU), Perceived usefulness (PU) and subjective norm (SN) [15]. Leyla Dogruelasuggested that new – media purposes are elaborated to describe the user's entertainment to elderly people. They are supposed to accept the new

technology based on the TAM structure. Elderly people planned to analyze the two concepts to use the 3D pictures for 50 age people and they work out the evaluation done for the computer game for the same age people. It made them accept the entertainment media technology for their mind-set to be involved in the enjoyment [16].

Senior Technology Acceptance Model (STAM) was developed to finalize with 1012 seniors aged 55 by the Hong Kong Chinese people. These people accept the gerontechnology based on STAM to gain the health information and features of theories of the age group. It observed 68% of different can be identified with the term gerontechnology. It assists to prognosticate the state of the age, gender, instruction, health details, the status of the diseases, number of affected technologies, age-related health care [17]. In 2018, wearable devices are implemented with traditional Chinese medicine (TCM) for detecting the elderly health condition. They used the diagnosis instrument to obtain all the information on the diseases or reports of the patients. Also, researchers proposed the Analysis of Variance (ANOVA) for changing the adoption of the changes between the people's health statement. It diagnoses the status of elderly people before and after the usage of wearable devices and without changing the information for detecting pulse duration [18].

Nelson and Dannefer demonstrated that heterogeneity is improved high when the age increased, and potential is contrasting for older adults. TAM analyzed that the elder's features are similar in their strength and convergence [19]. In [20] [21], devices are estimated to the adults for the hearing and vision disabilities in the text-based or graphical images. It is determined by using the mobile phone, digital camera, CD, MP3 player, alarms, pen drive etc to be supported for the elderly people as wearable devices. Ling (2013) [22] proposed an approach that the elderly people cannot find the exact movement and very slow in finding the activities such as writing a stylish letter, opening small targets, touching the small button. It is due to the diminishing of the touch thoughtfulness and conscious mental activity. In [23], about 50, 60, 70 age people compared with younger people. Elderly 50 age gave the result as 40%, 60 age peoples gave 40%, 70 age gave 20% and the younger age gave 57% engaged in the sampling rate from all people in Germany. They detect elderly women have more than males, whereas n = 125, female results as 54% and male results as 46% as per the observation.

Pan and Werner initiated that the older people from high-level status with higher education are dependent on the gerontechnology and suggest male candidates use better than the female candidate. In 2010 - 2011, aged people evolved PEOU, PU and AT for the observation of attitudes and lifestyles of the genders. It is also related to TAM to get the anxiety and flexibility of the action and much recent technology used for elderly people all over the world [24]. The wearing tags are used to recollect user acceptance in the environment for elderly people's care. They must use the 2 – dimensional systems with a better rate of 0.5m and 1 m with accurate “user Acceptance”. The wearable devices measure the elderly care by the system such accuracy (0.5 to 1 m), complexity during installing (less than 1 h), Not spread quickly, coverage area with 90m, sampling interval (0.5s) and the system is always accepted with availability [25].

Thota, C presented an efficient and secure centralized architecture for end to end integration of IoT based healthcare method deployed in Cloud environment. Sensor data is collected for health purposes, and the sensor data is safely communicated to near-edge devices. Finally, gadgets send data to the cloud, where it may be accessed by healthcare experts at any time. The major goal of this project is to ensure that all devices' authentication and authorization are safe [33]. Since then, sensors have become increasingly important in a wide range of applications, including environmental monitoring, transportation, smart city applications, and healthcare applications, among others. Wearable medical devices with sensors, in particular, are critical for collecting detailed health data. These sensors are constantly creating massive amounts of data, which is referred to as Big Data. Proposed a secure Industrial Internet of Things (IoT) architecture for health care applications to store and handle scalable sensor data (big data). Sensor medical devices are attached to the patient's body to collect clinical data. When the respiratory rate, heart rate, blood pressure,

body temperature, or blood sugar levels rise over normal, the devices send a clinically significant alarm message to the doctor via a wireless network. To protect large data in Industry 4.0, the suggested system employs a key management security method [34].

III. HYPOTHESIS AND RESEARCH MODEL

3.1 Research Model

Significant factors relating to the influence of smart wearable medical devices in users were developed. The proposed SWMDM model which assists the user in monitoring their health, the SWMDM model which consists of Techno Fear, Socioeconomic Characteristics, Self- potency and Gerontology concern as other parameters used to study users behaviour on wearable devices. Table 2 describes the definition of parameters which is used in the hypothesis for the analysis of adoption of proposed models in elder adults.

Table 2 Definition of Parameters

Parameters	Definition
Gerontology Concern	The state at which the person concerns that he/she is getting older.
Socioeconomic Characteristics	It describes the background, education and the current profession.
Self-Potency	It states the self-potential of a person to make an impression.
Techno Fear	The person's state of mind to adapt to the technology.
Technical Expertise	Advice from the technical expert on some technology.
Perceived Usefulness	State at which a person thinks that using a particular technology would increase his or her activities.
Behaviour Objective	The suitability of a particular smart device which has adverse behaviour.
Perceived ease of use	The limit at which a person would think that using the technology would free his or her efforts.

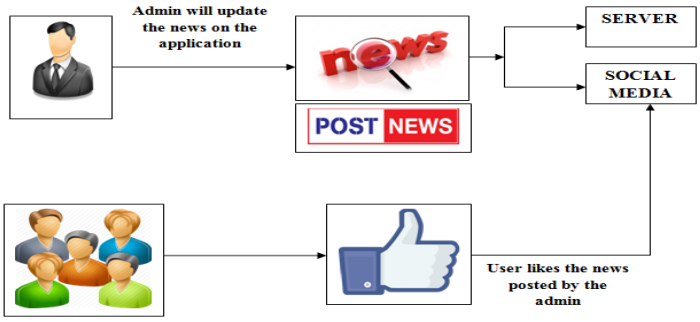


Fig. 2 SystemArchitecture

The figure 2 represents thesystemarchitecture, here the admin updates the news on the application, that will be transferred in the sever and the social media, By this the user will be able to view the news from various applications and gives their feedback on the news which is been collected by the server and will be for the statistics. It does, however, generate a vast number of rules and does not ensurethe efficacy or worth of the knowledge produced [35].

3.2 Research Hypothesis

Information security and privacy are becoming more crucial challenges in the healthcare industry. The use of digital patient records based on legislation, provider consolidation, and the increased need to exchange information among patients and providers are all examples of improved information security. Big data in health care is expected to enhance patient outcomes, forecast epidemic outbreaks, provide significant insights, prevent diseases, lower health-care costs, and improve quality-of-life analysis [36]. However, a trustworthy big data environment is ensured by the big data analytics-based cybersecurity architecture for security and privacy across health-care apps. Furthermore, electronic health records (EHR) may be shared by multiple users in order to improve the quality of health-care services. This raises serious privacy concerns that must be resolved before the EHR can be used. This architecture, which includes numerous technical methods and environmental controls, has been proved to be sufficient for appropriately addressing typical network security risks. Gao et al. (2014) [26] suggested in earlier studies that the relationship between the perceived usefulness (PU) which has been affected by the perceived ease of use (PEOU). The initial

model proposed by Davis in which he states that the perceived usefulness (PU) is affected by the perceived ease of use (PEOU). More realistic findings state that perceived usefulness is positively influenced by perceived ease of use. For instance, Van der Heijden et al. (2003) [27] came with the conclusion that, in environmental conditions the relationship between the perceived ease of use and perceived usef embraces true. ulness embraces true.

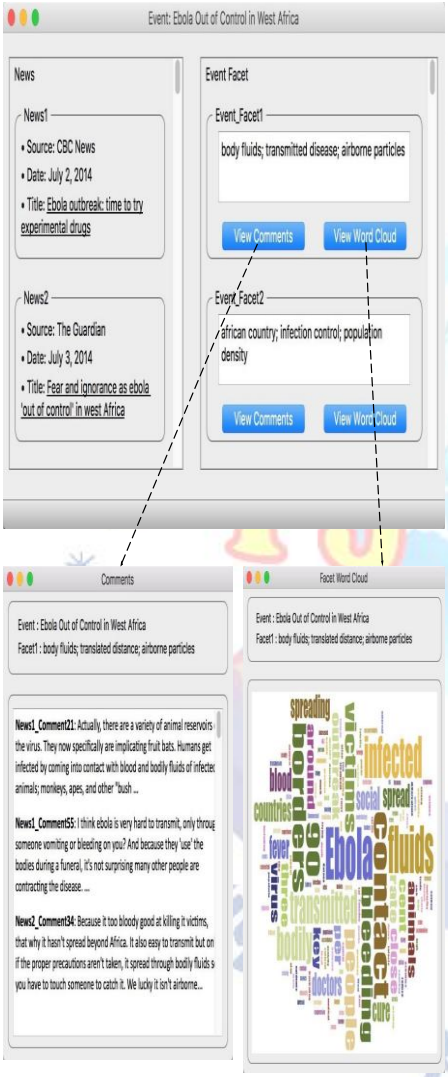


Fig. 3 Layers of the internal components.

Figure 3 describes proposed Layers of the internal componenets which has different parameters such as Socioeconomic Characteristics, Techno Fear, Technical Expertise, Gerontology Concern, Self-potency, and Behaviour Objective. These parameters are used to analyse how effective is the proposed model on the adaptation of smart wearable devices on elderly adults.

Application Layer: It handles all the connection between the administration and the users. This acts as an upper layer of all the layers and gives access to the other layers.

Logical Layer:

In this layer there are three types, Query Cache, Parser, Optimizer.

Query Cache: This will collect all the details of the user and storing the relevant retrieved records. So that if the client searches an information, it automatically fetches the data with less time consumption.

Parser: The parser is used to compiles the inputs in the form of tokens or command and sends the information or the command to the other layers.

Optimizer: These are a set of algorithms where is it user defined which is used to maximize the efficiency.

Techno Fear (TF)

Techno fear states the factor at which an elder user using a technology for his or her life support gets some fear of using some technology. Smart wearable devices assist elderly people who are reliable on these technologies but that fact to consider before these devices assisting people is whether the elderly people who have been monitored their health through these devices have fear of using the technology or having no fear of using the technology. In order to explain the in-depth analysis of how it affects the model is proposed through a hypothesis.

H11: TF has a positive effect on elder users' approach towards SWMDM.

Technical Expertise (TE)

Assisting the elderly adults using a technology related device will lead to a concern whether it is safe to use on the people. Technical expertise is the key factor influencing the act of knowing and analysing the technology to adults for monitoring their health. Here the expert will analyse the fact that the people who have been monitored are suitable for using the smart wearable devices in a stipulated time period. Based on the suggestion of the technical expert, the device is used on the elderly adult for assisting. We are proposing a hypothesis based on the technical expertise which is important in monitoring healthcare.

H12: TE has a positive effect on elder users' approach towards SWMDM.

3.3 Observance Study

This study has been observed by our proposed research model through the use of SWMDM in India.

3.3.1Instrumentation Development

The analysed instrumental measures from earlier research were utilised as the background for this study. By reviewing the previous studies, we could include some of the key parameters. For the betterment of the observed study, small modification of the quaternaries was made. Our study related questions were obtained from earlier studies developed by authors in [29] and altered to fit the methodology of using smart wearable devices. English is used as a language to develop the questionnaires; 30 amounts of items were used in the questionnaires. In accumulation, a six-point scale, with 1 giving negative of the scale (disagree strongly) and 6 giving positive of the scale (agree strongly), is used for participants' examination to get responses to the survey with questionnaires.

3.3.2Samples Data

Data used in this study was gathered through sheet-based questionnaires from 10th Jan to 20th Jan 2020 in the urban areas in the largest city in south India. Voluntarily participation of the people was requested. The purpose and the main objective of the survey were explained to the people. The intention of the study is mainly to focus on the aged adults. The participants were told that their information in the questionnaires would be held as private and the results would be given as aggregate. Smart wearable device experience of the people is considered to be the key objective of the survey. 250 questionnaires were collected from the people, among which 220 of the questionnaires are considered for the evaluation as it was a valid. Participants among the survey included 120 were male and 100 were female. The participants in the survey are over the age of 60.

Table 3 shows the statistical measures of each constructs in the proposed model

EventID	Title	#comments
1	Sochi terrorist attack	409
2	Malaysia Airlines disappearance	496
3	Kunming station massacre	157
4	New South Wales bush fires	84
5	Taiwan police evict student protesters	150
6	UK Cameron calls for action on superbug threat	209
7	Africa ebola out of control in West Africa	166
8	Ebola worker infected at WHO laboratory	239
9	ISIS in Iraq	252
10	Iwatch publish	189
11	Shut down of Malaysia Airlines MH17	138
12	The release of iPhone 6	215
13	US second ebola patient Nancy Writebol	155
14	BREMERTON Teen killing 6-year-old girlarrested	167
15	Boston Marathon bomber sister arrested	79
16	Beirut bombings attack	153
17	Apple and IBM team up	146
18	IOS8 released	161
19	Africa poachers killed elephants	163
20	ISIS executes David Haines	160
21	Rebels killed dozens in attack on refugees	162
22	Kent brantly cured Ebola patient	132
23	After Russian hacking	267
24	Japan whaling	134
25	Top doctor becomes latest Ebola victim	129
26	Air algerie plane crash	238
27	Bitcoin Mt Gox	442
28	California father accused killing family	153
29	UN host summit to end child brides	159
30	Great white shark choked by sea lion	115
31	New species of colorful monkey	178

Table 3 Statistical Measures of Each Items in SWMDM model

3.4 Hypothesis Model Measurement

The proposed hypothesis model quality is measured by validity of content, construct reliability and validity of discriminant which is mentioned by Bagozzi et al. (1979) [30]. To make sure that the validity of the content produced in the model, examination of

the questionnaire with 5 researchers with expertise in the field of information system was gathered on Dec 2019. The questionnaire given in the model was found to be efficient and was correctly understood by all the researchers.

Moreover, for proving the validity and reliability of every model's construct, coefficient alpha is calculated which states the reliability of each and every construct. Coefficient alpha is the analysis of the inner consistency of the model, it's reflected to be a degree of reliability scale. In our model, the coefficient alpha value was in range from 0.624 to 0.966. Robinson et al. (1991) [31] stated that coefficient value of 0.6 is considered to be a tolerable limit for coefficient alpha for experimental research. All the value in our research model was above 0.61. Subsequently, the scales in the model were satisfactory to stay.

Assessment of validity of convergent is carried out through composite reliability and average variance extracted (AVE). Bagozzi and Yi (2012) [32] stated that the given criteria used for measurements such as factor loading for every item in the model should surpass 0.5, the composite reliability should surpass 0.7 and AVE of every construct should surpass 0.5. Table 1 shows that all values of the measurement items are in standard range. Validity of discriminant measurement is shown in Table 4 that is used for analysing the results of the model; the gathered variances by the constructs are greater than the aligned correlation among variables. Constructs models' facts exposed were practically divergent. We attained better results for validity of convergent and discriminant, test result of the model's measurement was noble. Table 5 shows the Coefficient of Path based on Hypothesis.

Table 4 Validity of Discriminant

Variables	GC	PU	BO	SC	AT	SP	TF	TE	PEOU
GC	0.9								
PU	0.6	0.8							
BO	0.7	0.6	0.7						
SC	-0.2	-0.1	-0.2	0.6					
AT	0.4	0.5	0.6	0.5	0.5				
SP	0.1	0	0.5	0.6	0.1	0.7			
TF	0.2	0.	0.7	0.2	-0.	0.3	0.8		
		4			1				
	0.5	0.4	0.4	0.4	0.2	0.6	-0.1	0.6	
P	0.6	0.5	0.8	0.5	0.3	0.2	0.1	0.8	0.7

3.5 Testing of Hypothesis and Structural Model

WarpPLS software is used to test the model and the results are shown in Table 6. Nine (H1, H3, H4, H5, H6, H8, H9, H11, H12) out of twelve research hypotheses were considerably supported. Based on the results, Perceived Usefulness, Perceived ease of Use Behaviour Objective, Socioeconomic Characteristics, Technical Expertise were known to be statistically important effect on users' attitude towards SWMDM, whereas Gerontology Concern, Techno Fear, Self-Potency did not show important impact on users' behaviour of using SWMDM.

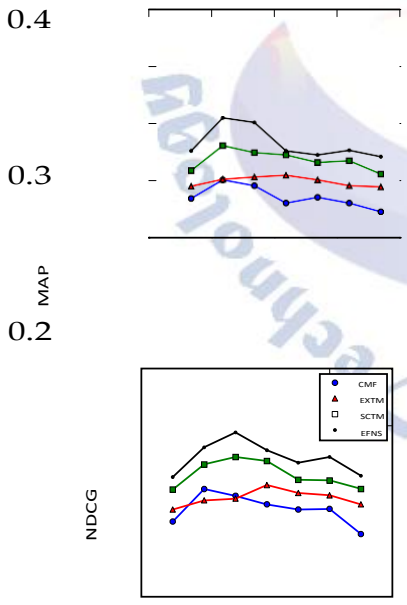
Hypothesis	Path's Coefficient	Analysis of Hypothesis
H1: GC to PU	0.698	Supported
H2: SWDB to AT	-0.095	Not Supported
H3: GC to AT	0.289	Supported
H4: PU to BO	0.356	Supported
H5: SP to BO	0.456	Supported
H6: TF to AT	0.564	Supported
H7: GC to BO	-0.156	Not Supported
H8: PEOU to AT	0.542	Supported
H9: SC to AT	0.4	Supported
H10: SC to BT	0.052	Not Supported
H11: TF to AT	0.256	Supported
H12: TE to AT	0.343	Supported

Table 5 ANOVA of the acceptance of smart wearable devices

IV. RESULTS AND DISCUSSIONS

The conclusion of the study made by the proposed SWMDM on the adoption of user wearable devices on the elderly adult were made through the Hypothesis and questionnaires were developed and the model was analysed based the questionnaires answered from elderly adults in the southern city of India. This study was evaluated according to the literature review from the previous models on the adoption of wearable devices on users. There were various factors influencing the adoption of smart wearable devices on the elder users, which includes factors such as socioeconomic characteristics, techno fear, technical expertise, gerontology concern. On the other hand, hypothesis was developed based on the factors considered for evaluation in smart wearable devices. Inference from the result indicates, found that 9 proposed hypotheses were supported. Self-potency (0.454), PEOU (0.542), Gerontology Concern (0.698) and Techno Fear (0.564) explains 54.5 percent of the perceived variance in users' attitude toward SWMDM. Substantially positive impact on users' attitude towards SWMDM is seen by having both PU and PEOU.

Fig. 4 Proposed SWMDM Models Statistical Analysis



	CMF		EXTM		SCTM		EFNS	
	MAP	NDCG	MAP	NDCG	MAP	NDCG	MAP	NDCG
eventfacet oriented	0.142	0.172	0.151	0.191	—	—	0.212	0.299
news general	0.183	0.232	0.188	0.243	0.173	0.254	0.281	0.311
news specific	0.113	0.179	0.132	0.181	0.261	0.296	0.282	0.307
Avg.	0.146	0.194	0.157	0.205	0.213	0.275	**0.258**	**0.305**

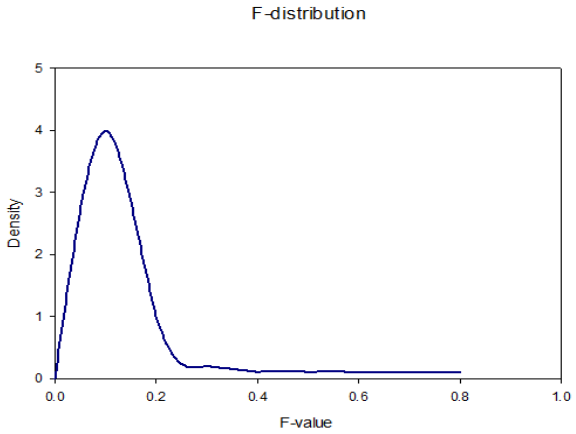


Fig. 5 F-distribution based on Hypothesis

Statistical Measurements of every item in the SWMDM model is shown in figure 4. Every scales item internal consistency is measured using composite reliability, the construct variance amount is calculated using average variance extracted (AVE) and variance score of every construct observed is measured using Coefficient alpha. Average variance extracted is higher in Techno Fear, composite reliability is greater in Gerontology Concern and Technical Expertise shows higher rate of coefficient alpha. Figure 5 shows the F-distribution of the Anova test conducted on the proposed smart wearable device model, the F-value or score is calculated which determines the probability distribution function. F- Value is lower at higher density which states that the results obtained from the hypothesis are significant.

V. FUTURE SCOPE AND CONCLUSION

Thus, in this project, we conclude that through this system we can get the latest news of various categories on time and we can avoid manual errors as well as time consumption in segregation of data by performing this task through automated process. Also, we can not only get the latest news only when we launch our application but also through emails/SMS by using the Automated Alert System so that user will save time and gets his/her subscribed news on time.

REFERENCES

1. T. Rosenstiel, Twitter and the News: How People use the Social Network to Learn About the World. Arlington, VA, USA: American Press Institute, 2015.
2. Z. Ma, A. Sun, Q. Yuan, and G. Cong, "Tagging your tweets:Aprobabilistic modeling of hashtag annotation in

twitter," in Proc. 23rd ACMInt. Conf. Conf. Inform. Knowl.Manage., 2014, pp. 999–1008.
3. T.-A. Hoang-Vu, A. Bessa, L. Barbosa, and J. Freire, "Bridging vocabularies to link tweets and news," in Proc. Int. Workshop Web Databases, 2014.
4. Y. Gong, Q. Zhang, and X. Huang, "Hashtag recommendation using dirichlet process mixture models incorporating types of hashtags," in Proc. Empirical Methods Natural Language Process., 2015, pp. 401–410.
5. Y. Gong and Q. Zhang, "Hashtag recommendation using attention- based convolutional neural network," in Proc. 25th Int. Joint Conf. Artif. Intell., 2016, pp. 2782–2788.
6. B. Shi, G. Ifrim, and N. Hurley, "Learning-to-rank for real-time high-precision hashtag recommendation for streaming news," in Proc. 25th Int. Conf. World Wide Web, 2016, pp. 1191–1202.
7. K. S. Hasan and V. Ng, "Automatic key phrase extraction: A survey of the state of the art," in Proc. 52nd Annu. Meeting Assoc. Compute. Linguistics, 2014, pp. 1262–1273.
8. T.-Y. Liu, "Learning to rank for information retrieval," Foundations Trends Inform. Retrieval, vol. 3, pp. 225–331, 2009.
9. R. Dovgopol and M. Nohelty, "Twitter hash tag recommendation," CoRR, vol. abs/1502.00094, 2015, http://arxiv.org/abs/1502.00094
10. A. Mazzia and J. Juett, "Suggesting hashtags on twitter," EECS 545 Project, 2011, http://www-personal.umich.edu/~amazzia/pubs/ 545-final.pdf.
11. B. Shi, W. Lam, L. Bing, and Y. Xu, "Detecting common discussion topics across culture from news reader comments," in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016, pp. 676–685.
12. T. Bansal, M. Das, and C. Bhattacharyya, "Content driven user profiling for comment-worthy recommendations of news and blog articles," in Proceedings of the 9th ACM Conference on Recommender Systems, 2015, pp. 195–202.
13. A. Aker, E. Kurtic, A. Balamurali, M. Paramita, E. Barker, M. Hepple, and R. Gaizauskas, "A graph-based approach to topic clustering for online comments to news," in Proceedings of the 38th European Conference on IR Research, 2016, pp. 15–29.
14. Q. Li, J. Wang, Y. P. Chen, and Z.Lin, "User comments for news recommendation in forum-based social media," Information Sciences, vol. 180, no. 24, pp. 4929–4939, 2010.
15. G. Rizos, S. Papadopoulos, and Y. Kompatsiaris, "Predicting news popularity by mining online discussions," in Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Companion Volume, 2016, pp. 737–742.
16. L. Jianghao, Z. Yongmei, Y. Aimin, and C.Jin, "Analysis of topic evolution on news comments based on word vectors," in International Conference on Cloud Computing and Security. Springer, 2016, pp. 464–475.
17. C. Wang, Y. Zhang, W. Jie, C. Sauer, and X. Yuan, "Hierarchical semantic representations of online news comments for emotion tagging using multiple information sources," in International Con- ference on Database Systems for Advanced Applications. Springer, 2017, pp. 121–136.
18. J. Grimmer, "A bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases," Political Analysis, vol. 18, no. 1, pp. 1–35, 2010.
19. Y. Fang, L. Si, N. Somasundaram, and Z. Yu, "Mining contrastive opinions on political texts using cross-perspective topic model," in Proceedings of the 5th ACM International Conference on Web Search and Data Mining, 2012, pp. 63–72.
20. I. Titov and R. McDonald, "Modeling online reviews with multi- grain topic models," in Proceedings of the 17th International Confer- ence on World Wide Web, 2008, pp. 111–120.
21. D. M. Blei and P. J. Moreno, "Topic segmentation with an as- pect hidden markov model," in Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2001, pp. 343–348.
22. L. Du, W. L. Buntine, and M. Johnson, "Topic segmentation with a structured topic model," in Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2013, pp. 190–200.
23. F. Shahnaz, M. W. Berry, V. P. Pauca, and R. J. Plemmons, "Document clustering using nonnegative matrix factorization," Information Processing & Management, vol. 42, no. 2, pp. 373–386, 2006
24. J. Kalyanam, A. Mantrach, D. Saez-Trumper, H. Vahabi, andG. Lanckriet, "Leveraging social context for modeling topic evolu- tion," in Proceedings of the 21st ACM SIGKDD International Confer- ence on Knowledge Discovery and Data Mining, 2015, pp. 517–526.
25. C. K. Vaca, A. Mantrach, A. Jaimes, and M. Saerens, "A time- based collective factorization for topic discovery and monitoring in news," in Proceedings of the 23rd International Conference on World Wide Web, 2014, pp. 527–538.
26. C.-J. Lin, "Projected gradient methods for nonnegative matrix factorization," Neural Computation, vol. 19, no. 10, pp. 2756–2779, 2007.J. Mairal, F. Bach, J.Ponce, and

G.Sapiro, "Online learning for matrix factorization and sparse coding," Journal of Machine Learning Research, vol. 11, pp. 19–60, 2010.

27. N. Srebro, J. Rennie, and T. S. Jaakkola, "Maximum-margin matrix factorization," in Advances in Neural Information Processing Systems, 2004, pp. 1329–1336.
28. M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and
29. R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," Computational Statistics & Data Analysis, vol. 52, no. 1, pp. 155–173, 2007.
30. G. Bouchard, D. Yin, and S. Guo, "Convex collective matrix factorization," in Proceedings of the 16th International Conference on Artificial Intelligence and Statistics, 2013, pp. 144–152.
31. P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," Journal of Machine Learning Research, vol. 5, pp. 1457– 1469, 2004.
32. H. Bang and J.-H. Lee, "Collective matrix factorization using tag embedding for effective recommender system," in Proceedings of the 17th International Symposium on Advanced Intelligent Systems, 2016, pp. 846–850.
33. N. Aletras, T. Baldwin, J. H. Lau, and M. Stevenson, "Evaluating topic representations for exploring document collections," Journal of the Association for Information Science and Technology, vol. 68, no. 1, pp. 154–167, 2015.
34. Q. Mei, X. Shen, and C. Zhai, "Automatic labeling of multinomial topic models," in Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2007, pp. 490– 499.