

*International Conference on Computer, Engineering, Law, Education and  
Management 2017*

---

**ICCELEM 2017**



*International Conference on Computer, Engineering, Law, Education and  
Management 2017*

---

# Volume 1

By  
ASDF International

Financially Sponsored By  
Association of Scientists, Developers and Faculties, India

Multiple Areas

28 – 29, September 2017  
The Westin Chosun Seoul,  
Seoul, South Korea

*Editor-in-Chief*  
**Dr. K Kokula Krishna Hari**

*Editors:*

Daniel James & Saikishore Elangovan

*Published by*

**Association of Scientists, Developers and Faculties**

Address: RMZ Millennia Business Park, Campus 4B, Phase II, 6<sup>th</sup> Floor, No. 143, Dr. MGR Salai, Kandanchavady, Perungudi, Chennai – 600 096, India.

Email: admin@asdf.org.in || www.asdf.org.in

**International Conference on Computer, Engineering, Law, Education and Management (ICCELEM 2017)**

**VOLUME 1**

Editor-in-Chief: **Dr. K Kokula Krishna Hari**

Editors: **Daniel James & Saikishore Elangovan**

Copyright © 2017 ICCELEM 2017 Organizers. All rights Reserved

This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the ICCELEM 2017 Organizers or the Publisher.

**Disclaimer:**

No responsibility is assumed by the ICCELEM 2017 Organizers/Publisher for any injury and/ or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products or ideas contained in the material herein. Contents, used in the papers and how it is submitted and approved by the contributors after changes in the formatting. Whilst every attempt made to ensure that all aspects of the paper are uniform in style, the ICCELEM 2017 Organizers, Publisher or the Editor(s) will not be responsible whatsoever for the accuracy, correctness or representation of any statements or documents presented in the papers.

ISBN-13: 978-81-933584-7-4

ISBN-10: 81-933584-7-3

## **PREFACE**

The International Conference on Computer, Engineering, Law, Education and Management (ICCELEM 2017)” held on 28 - 29<sup>th</sup> September 2017, in collaboration with Association of Scientists, Developers and Faculties (ASDF), an International body, at The Westin Chosun Seoul, Seoul, South Korea.

ICCELEM 2017 provides a chance for academic and Industry professionals to discuss the recent progress in the area of Computer, Engineering, Law, Education and Management. The outcome of the conference will trigger for the further related research and future technological improvement. This conference highlights the novel concepts and improvements related to the research and technology.

The technical committee consists of experts in the various course subfields helped to scrutinize the technical papers in various fields, support to maintain the quality level of the proceedings of conference which consist of the information of various advancements in the field of research and development globally and would act as a primary resource of researchers to gain knowledge in their relevant fields.

The constant support and encouragement from Dr. S. Prithiv Rajan, ASDF Global President, Dr. P. Anbuoli, ASDF International President helped a lot to conduct the conference and to publish the proceedings within a short span. I would like to express my deep appreciation and heartfelt thanks to the ASDF team members. Without them, the proceedings could not have been completed in a successful manner. I would like to express my sincere thanks to our management, student friends and colleagues for their involvement, interest, enthusiasm to bring this proceeding of the conference in a successful way.

**Dr. K Kokula Krishna Hari,**

Chief Editor cum Convener,

General Secretary, ASDF International, London, United Kingdom

# **Organizing Committee**

## **Conference Honorary Chair**

**Tai Hoon Kim**, Professor, South Korea

### *Chief Patron*

- **S Prithiv Rajan**, Managing Trustee, Techno Forum Research & Development Centre, India

### *Patrons*

- **R Sudhakar**, Supporting Trustee, Techno Forum Research & Development Centre, India

### *Convener*

- **P Anbuoli**, Trustee, Techno Forum Research & Development Centre, India

### *Organizing Secretaries*

- **E Saikishore**, Supporting Trustee, Techno Forum Research & Development Centre, India

### *Publications Chair & Thesis Chair*

- **Dr Daniel James**, London, United Kingdom

### **TECHNICAL REVIEWERS**

- A Amsavalli, Paavai Engineering College, Namakkal, India
- A Ayyasamy, Annamalai University, Chidambaram, India
- A C Shagar, Sethu Institute of Technology, India
- A Kavitha, Chettinad College of Engineering & Technology, Karur, India
- A Padma, Madurai Institute of Engineering and Technology, Madurai, India
- A S N Chakravarthy, JNTU Kakinada, India
- A Tamilarasi, Kongu Engineering College, Perundurai, India
- Abdelbasset Brahim, University of Granada, Spain
- Abdelnaser Omran, Universiti Utara Malaysia, Malaysia
- Abdul Aziz Hussin, Universiti Sains Malaysia, Malaysia
- Abdul Nawfar Bin Sadagatullah, Universiti Sains Malaysia, Malaysia
- Abhay Prabhakar Kulkarni, Director - IICMR, Pune

- Abhishek Bajpai, SRM University, Lucknow, India
- Abhishek Shukla, U.P.T.U. Lucknow, India
- Aede Hatib Musta'amal, Universiti Teknologi Malaysia, Malaysia
- Ahmed Mohammed Kamaruddeen, Universiti Utara Malaysia, Malaysia
- Ahmed Salem, Old Dominion University, United States of America
- Ali Berkol, Baskent University & Space and Defence Technologies (SDT), Turkey
- Alphin M S, SSN College of Engineering, Chennai, India
- Alwardoss Velayutham Raviprakash, Pondicherry Engineering College, Pondicherry, India
- Anand Nayyar, KCL Institute of Management and Technology, Punjab
- Anbuezhayan M, Valliammai Engineering College, Chennai, India
- Ang Miin Huey, Universiti Sains Malaysia, Malaysia
- Anirban Mitra, VITAM Berhampur, Odisha, India
- Ariffin Abdul Mutalib, Universiti Utara Malaysia, Malaysia
- Arniza Ghazali, Universiti Sains Malaysia, Malaysia
- Arumugam Raman, Universiti Utara Malaysia, Malaysia
- Aruna Anil Deoskar, IICMR, Pune, India
- Asha Ambhaikar, Rungta College of Engineering & Technology, Bhilai, India
- Ashish Chaurasia, RGPV, Bhopal, Madhya Pradesh
- Asrulnizam Bin Abd Manaf, Universiti Sains Malaysia, Malaysia
- Ata Elahi, Southern Connecticut State University, USA
- Aziah Daud, Universiti Sains Malaysia, Malaysia
- B Paramasivan, National College of Engineering, Tirunelveli, India
- Badruddin A. Rahman, Universiti Utara Malaysia, Malaysia
- Balachandran Ruthramurthy, Multimedia University, Malaysia
- Balasubramanie Palanisamy, Professor & Head, Kongu Engineering College, India
- Brahim Abdelbasset, University of Granada, Spain
- C Poongodi, Bannari Amman Institute of Technology, Sathyamangalam, India
- Chandrasekaran Subramaniam, Professor & Dean, Anna University, India
- Chitra Krishnan, VIT University, Chennai, India
- Chokri Ben Amar, University of Sfax, Tunisia
- Choo Ling Suan, Universiti Utara Malaysia, Malaysia
- Cristian-Gyozo Haba, Technical University of Iasi, Romania
- D Deepa, Bannari Amman Institute of Technology, Sathyamangalam, India

- D Gracia Nirmala Rani, Thiagarajar College of Engineering, Madurai, Tamil Nadu
- D Sheela, Tagore Engineering College, Chennai, India
- Daniel James, Senior Researcher, United Kingdom
- David Rathnaraj Jebamani, Sri Ramakrishna Engineering College, India
- Deepali Sawai, Director - MCA, University of Pune ( Savitribai Phule Pune University ),  
India
- Dewi Nasien, Universiti Teknologi Malaysia, Malaysia
- Doug Witten, Oakland University, Rochester, United States of America
- Dzati Athiar Ramli, Universiti Sains Malaysia, Malaysia
- E Bhaskaran, Government of Tamilnadu, Chennai, India
- Fadhilah Mat Yamin, Universiti Utara Malaysia, Malaysia
- G A Sathish Kumar, Sri Venkateswara College of Engineering, India
- G Arunkumar, Saveetha University, Chennai, India
- G Ganesan, Adikavi Nannaya University, India
- G Subbaraju, Shri Vishnu Engineering College for Women, India
- Ganesan Kanagaraj, Thiagarajar College of Engineering, Madurai, Tamil Nadu
- Geetha G, Jerusalem College of Engineering, Chennai, India
- Geetha V, Pondicherry Engineering College, Pondicherry, India
- Guobiao Yang, Tongji University, China
- Hanumantha Reddy T, RYM Engineering College, Bellary, India
- Hardeep Singh Saini, Indo Global College of Engineering, Mohali, Punjab
- Hareesh N Ramanathan, Toc H Institute of Science and Technology, India
- Hari Mohan Pandey, Amity University, Noida, India
- Helena Karsten, Abo Akademi University, Finland
- Hidayani Binti Jaafar, Universiti Malaysia Kelantan, Malaysia
- Itebeddine GHORBEL, INSERM, France
- J Baskaran, Adhiparasakthi Engineering College, Melmaruvathur, India
- J Karthikeyan, Anna University, Chennai, India
- J Sadhik Basha, International Maritime College, Oman
- Jebaraj S, Universiti Teknologi PETRONAS (UTP), Malaysia
- Jia Uddin, International Islamic University Chittagong, Bangladesh
- Jinnah Sheik Mohamed M, National College of Engineering, Tirunelveli, India
- John Augustine P, Sri Eshwar College of Engineering, Coimbatore, India
- Julie Juliewatty Mohamed, Universiti Malaysia Kelantan, Malaysia



- K Latha, Anna University, Chennai, India
- K Mohamed Bak, Ilahia School of Science and Technology, India
- K Nirmalkumar, Kongu Engineering College, Perundurai, India
- K P Kannan, Bannari Amman Institute of Technology, Sathyamangalam, India
- K Parmasivam, K S R College of Engineering, Thiruchengode, India
- K Senthilkumar, Erode Sengunthar Engineering College, Erode, India
- K Suriyan, Bharathiyar University, India
- K Thamizhmaran, Annamalai University, Chidambaram, India
- K Thiruppathi, Valliammai Engineering College, India
- K Vijayaraja, PB College of Engineering, Chennai, India
- Kamal Imran Mohd Sharif, Universiti Utara Malaysia, Malaysia
- Kannan G R, PSNA College of Engineering and Technology, Dindigul, India
- Kathiravan S, Kalaignar Karunanidhi Institute of Technology, Coimbatore, India
- Khairul Anuar Mohammad Shah, Universiti Sains Malaysia, Malaysia
- Khurram Saleem Alimgeer, COMSATS Institute of Information Technology, Islamabad
- Kokula Krishna Hari Kunasekaran, Chief Scientist, Techno Forum Research and Development Center, India
- Konguvel Elango, Dhanalakshmi Srinivasan College of Engineering, Coimbatore
- Krishnan J, Annamalai University, Chidambaram, India
- Kumarathan N, Sri Venkateswara College of Engineering, India
- L Ashok Kumar, PSG College of Technology, Coimbatore, India
- Laila Khedher, University of Granada, Spain
- Lakshmanan Thangavelu, SA College of Engineering, Chennai, India
- M Ayaz Ahmad, University of Tabuk, Saudi Arabia
- M Chandrasekaran, Government College of Engineering, Bargur, India
- M K Kavitha Devi, Thiagarajar College of Engineering, Madurai, Tamil Nadu
- M Karthikeyan, Knowledge Institute of Technology, India
- M Shanmugapriya, SSN College of Engineering, Chennai, India
- M Thangamani, Kongu Engineering College, India
- M Venkatachalam, RVS Technical Campus - Coimbatore, India
- M Vimalan, Thirumalai Engineering College, Kanchipuram, India
- Malathi R, Annamalai University, Chidambaram, India
- Mansoor Zoveidavianpoor, Universiti Teknologi Malaysia, Malaysia
- Manvender Kaur Chahal, Universiti Utara Malaysia, Malaysia

- Mariem Mahfoudh, MIPS, France
- Marinah Binti Othman, Universiti Sains Islam Malaysia, Malaysia
- Mathivannan Jaganathan, Universiti Utara Malaysia, Malaysia
- Md Haider Ali Biswas, Khulna University, Khulna, Bangladesh
- Md Nur Alam, Pabna university of Science & Technology, Bangladesh
- Mehdi Asadi, IAU (Islamic Azad University), Iran
- Mohamed Moussaoui, ENSA of Tangier Abdelmalek Essaadi University, Morocco
- Mohamed Saber Mohamed Gad, National Research Center, Egypt
- Mohammad Ayaz Ahmad, University of Tabuk, Saudi Arabia
- Mohammed Ali Hussain, KL University, India
- Mohan Awasthy, Chhattisgarh Swami Vivekanand Technical University, Bilai, Chhattisgarh
- Mohd Hanim Osman, Universiti Teknologi Malaysia, Malaysia
- Mohd Hashim Siti Z, Universiti Teknologi Malaysia, Malaysia
- Mohd Helmy Abd Wahab, Universiti Tun Hussein Onn, Malaysia
- Mohd Murtadha Mohamad, Universiti Teknologi Malaysia, Malaysia
- Mohd Zulkifli Bin Mohd Yunus, Universiti Teknologi Malaysia, Malaysia
- Moniruzzaman Bhuiyan, University of Northumbria, United Kingdom
- Mora Veera Madhava Rao, Osmania University, India
- Muhammad Iqbal Ahmad, Universiti Malaysia Kelantan, Malaysia
- Muhammad Javed, Cornell University, United States of America
- Mukesh D Patil, Ramrao Adik Institute of Technology, India
- Mukesh Negi, TechMahindra Ltd, India
- N Karthikeyan, SNS College of Engineering, Coimbatore, India
- N Malmurugan, Mahendra Group of Institutions, India
- N Meenakshi Sundaram, PSG College of Technology, Coimbatore, India
- N Rajesh Jesudoss Hynes, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India
- N Senthilnathan, Kongu Engineering College, Perundurai, India
- N Shanthi, Nandha Engineering College, Erode, India
- N Suthanthira Vanitha, Knowledge Institute of Technology, India
- Nasrul Humaimi Mahmood, Universiti Teknologi Malaysia, Malaysia
- Nida Iqbal, Universiti Teknologi Malaysia, Malaysia
- Nithya Kalyani S, K S R College of Engineering, Thiruchengode, India

- Nor Muzlifah Mahyuddin, Universiti Sains Malaysia, Malaysia
- Norma Binti Alias, Universiti Teknologi Malaysia, Malaysia
- O L Shanmugasundaram, K S R College of Engineering, Thiruchengode, India
- P Dhanasekaran, Erode Sengunthar Engineering College, Erode, India
- P Ganesh Kumar, K. L. N. College of Information Technology, Madurai, India
- P Kumar, K S R College of Engineering, Thiruchengode, India
- P Ramasamy, Sri Balaji Chockalingam Engineering College, India
- P Raviraj, Kalaignar Karunanidhi Institute of Technology, Coimbatore, India
- P Sengottuvelan, Bannari Amman Institute of Technology, Sathyamangalam, India
- P Shunmuga Perumal, Anna University, Chennai, India
- P Sivakumar, K S R College of Engineering, Thiruchengode, India
- P Sudhakar, M Kumarasamy College of Engineering, Karur, India
- P Tamizhselvan, Bharathiyar University, India
- P Thamilarasu, Paavai Engineering College, Namakkal, India
- Pasupuleti Visweswara Rao, Universiti Malaysia Kelantan, Malaysia
- Pethuru Raj, IBM Research, India
- Qais Faryadi, USIM: Universiti Sains Islam Malaysia, Malaysia
- R Ashokan, Kongunadu College of Engineering and Technology, India
- R Dhanasekaran, Syed Ammal Engineering College, Ramanathapuram, India
- R Muthukumar, Shree Venkateshwara Hi-Tech Engineering College, India
- R Nallusamy, Principal, Nandha college of Technology, Erode, India
- R Ragupathy, Annamalai University, Chidambaram, India
- R Sudhakar, Dr. Mahalingam College of Engineering and Technology, India
- R Suguna, SKR Engineering College, Chennai, India
- R Sundareswaran, SSN College of Engineering, Chennai, India
- Radzi Ismail, Universiti Sains Malaysia, Malaysia
- Raghvendra Kumar, LNCT College, Jabalpur
- Rajesh Deshmukh, Shri Shankaracharya Institute of Professional Management and Technology, Raipur
- Rathika P, V V College of Engineering, Tirunelveli, India
- Rathinam Maheswaran, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India
- Ravindra W Gaikwad, Pravara Rural Engineering College, Loni
- Razauden Mohamed Zulkifli, Universiti Teknologi Malaysia, Malaysia
- Reza Gharoie Ahangar, University of North Texas, USA

- Roesnita Ismail, USIM: Universiti Sains Islam Malaysia, Malaysia
- Rohaizah Saad, Universiti Utara Malaysia, Malaysia
- Roselina Binti Sallehuddin, Universiti Teknologi Malaysia, Malaysia
- Ruba Soundar K, P. S. R. Engineering College, Sivakasi, India
- S Albert Alexander, Kongu Engineering College, Perundurai, India
- S Anand, V V College of Engineering, Tirunelveli, India
- S Appavu @ Balamurugan, K. L. N. College of Information Technology, Madurai, India
- S Balaji, Jain University, India
- S Balamuralitharan, SRM University, Chennai, India
- S Balamurugan, Kalaignar Karunanidhi Institute of Technology, Coimbatore, India
- S Geetha, VIT University, Chennai, India
- S Jaganathan, Dr. N. G. P. Institute of Technology, Coimbatore, India
- S Karthik, SNS College of Technology, India
- S Natarajan, Karpagam College of Engineering, Coimbatore, India
- S Nithyanandam, PRIST University, India
- S Poorani, Karpagam University, Coimbatore, India
- S Prakash, Nehru Colleges, Coimbatore, India
- S R Kumbhar, Rajarambapu Institute of Technology, India
- S Rajkumar, University College of Engineering Ariyalur, India
- S Ramesh, Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College, India
- S Selvaperumal, Syed Ammal Engineering College, Ramanathapuram, India
- S Selvi, Institute of Road and Transport Technology, India
- S Senthamarai Kannan, Kalasalingam University, India
- S Senthilkumar, Sri Shakthi Institute of Engineering and Technology, Coimbatore, India
- S Shahil Kirupavathy, Velammal Engineering College, Chennai, India
- S Vengataasalam, Kongu Engineering College, Perundurai, India
- Samuel Charles, Dhanalakshmi Srinivasan College of Engineering, Coimbatore, India
- Sangeetha R G, VIT University, Chennai, India
- Sanjay Singhal, Founder, Strategizers, India
- Sanjeevikumar Padmanaban, Ohm Technologies, India
- Saratha Sathasivam, Universiti Sains Malaysia, Malaysia
- Sarina Sulaiman, Universiti Teknologi Malaysia, Malaysia
- Sathish Kumar Nagarajan, Sri Ramakrishna Engineering College, Coimbatore, India
- Sathishbabu S, Annamalai University, Chidambaram, India

- Seddik Hassene, ENSIT, Tunisia
- Selvakumar Manickam, Universiti Sains Malaysia, Malaysia
- Shamshuritawati Sharif, Universiti Utara Malaysia, Malaysia
- Shankar S, Kongu Engineering College, Perundurai, India
- Shazida Jan Mohd Khan, Universiti Utara Malaysia, Malaysia
- Sheikh Abdul Rezan, Universiti Sains Malaysia, Malaysia
- Shilpa Bhalerao, Acropolis Institute of Technology and Research, Indore, India
- Singaravel G, K. S. R. College of Engineering, India
- Sivakumar Ramakrishnan, Universiti Sains Malaysia, Malaysia
- Smriti Agrawal, Chiatanya Bharathi Institute of Technology, Hyderabad
- Somasundaram Sankaralingam, Coimbatore Institute of Technology, India
- Sri Devi Ravana, University of Malaya, Malaysia
- Subash Chandra Bose Jeganathan, Professional Group of Institutions, India
- Subramaniam Ganesan, Oakland University, Rochester, United States of America
- Suganthi Appalasamy, Universiti Malaysia Kelantan, Malaysia
- Sundar Ganesh C S, PSG College of Technology, Coimbatore, India
- Sunil Chowdhary, Amity University, Noida, India
- Sunita Daniel, Amity University, Haryana
- Suresh Sagadevan, Indian Institute of Science, Bangalore, India
- Syed Sahal Nazli Alhady, Universiti Sains Malaysia, Malaysia
- T K P Rajagopal, Kathir College of Engineering, Coimbatore, India
- T Krishnakumar, Tagore Engineering College, Chennai, India
- T Ramayah, Universiti Sains Malaysia, Malaysia
- T Subbulakshmi, VIT University, Chennai, India
- T V P Sundararajan, Bannari Amman Institute of Technology, Sathyamangalam, India
- Tamilarasi Angamuthu, Kongu Engineering College, Perundurai, India
- Tom Kolan, IBM Research, Israel
- Uma N Dulhare, Muffkham Jah College of Engineering & Technology, Hyderabad, India
- Uvaraja V C, Bannari Amman Institute of Technology, Sathyamangalam, India
- V Akila, Pondicherry Engineering College, Pondicherry, India
- V C Sathish Gandhi, University College of Engineering Nagercoil, India
- V E Nethaji Mariappan, Sathyabama University, India
- V Mohanasundaram, Vivekanandha Institute of Engineering and Technology for Women, India

- V Ramesh, Mahatma Gandhi Institute of Technology, Hyderabad
- V Sathish, Bannari Amman Institute of Technology, Sathyamangalam, India
- V Vijayakumari, Sri Krishna College of Technology, Coimbatore, India
- Vaiyapuri Govindasamy, Pondicherry Engineering College, Pondicherry, India
- Veera Jyothi Badnal, Osmania University, India
- Veeraswamy Ammisetty, St. Ann's College of Engineering & Technology, India
- Venkatesh MP, Annamalai University, Chidambaram, India
- Vijayalakshmi V, Pondicherry Engineering College, Pondicherry, India
- Vijayan Gurusurthy Iyer, Entrepreneurship Development Institute of India
- Vikrant Bhateja, Shri Ramswaroop Memorial Group of Professional Colleges (SRMGPC), India
- Wan Hussain Wan Ishak, Universiti Utara Malaysia, Malaysia
- Wei Ping Loh, Universiti Sains Malaysia, Malaysia
- Yatay Sulaiman, Universiti Utara Malaysia, Malaysia
- Yerra Rama Mohana Rao, Dr. Pauls Engineering College, India
- Yongan Tang, Oakland University, Rochester, United States of America
- Yousef FARHAOUI, Moulay Ismail University, Morocco
- Yudi Fernando, Universiti Sains Malaysia, Malaysia
- Yu-N Cheah, Universiti Sains Malaysia, Malaysia
- Zahurin Samad, Universiti Sains Malaysia, Malaysia
- Zailan Siri, University of Malaya, Malaysia
- Zainuddin Bin Zakaria, Universiti Teknologi MARA, Dungun Campus, Terengganu
- Zamira Zamzuri, Universiti Kebangsaan Malaysia, Malaysia
- Zul Ariff Abdul Latiff, Universiti Malaysia Kelantan, Malaysia

# Table of Content

<b>Volume</b>	01	<b>ISBN</b>	978-81-933584-7-4
<b>Month</b>	September	<b>Year</b>	2017

International Conference on Computer, Engineering, Law, Education and Management 2017

<b>Title &amp; Authors</b>	<b>Pages</b>
<b>Embedding the Multiple Linear Regression Model to Monitor Student Performance in the Flexible Digital Learning Environment</b> <i>by Benilda Eleonor V Comendador</i>	pp01 – pp05
<b>Sentiment Analysis on Service Complaints using Sarcasm and Emoticon Detection</b> <i>by Ria A Sagum, Clowie Fel Y Bingco, J Wesley P Florentino</i>	pp06 – pp11
<b>Diagnostic App for Cats and Dogs Diseases using Neuro – Fuzzy Algorithm</b> <i>by Carlo G Inovero, Emmanuel Vincent Gratila, Jane Maelyne C Lopez</i>	pp12 – pp16
<b>Automatic Classification of Consumer Queries using LSTM</b> <i>by Long Cai, Kokula Krishna Hari K, Rajkumar Sugumaran</i>	pp17 – pp20



ISBN	978-81-933584-7-4
Website	www.iccelem.com
Received	30-May-2017
Article ID	ICCELEM001

VOL	01
eMail	iccelem@asdf.res.in
Accepted	25-July-2017
eAID	ICCELEM.2017.001

# Embedding the Multiple Linear Regression Model to Monitor Student Performance in the Flexible Digital Learning Environment

Benilda Eleonor V Comendador<sup>1</sup>

<sup>1</sup>Polytechnic University of the Philippines, Anonas St., Sta. Mesa, Manila, Philippines

**Abstract**—This study focused on providing a Decision Support System (DSS) that integrates Multiple Linear Regression (MLR) model to monitor student performance in the flexible digital learning environment. The author carried out series of experiments in order to evaluate the performance and usefulness of the generated models. MLR was adopted in the development of the Learning Analytics Decision Support System. The developed application predicts the performance of the university portal users which may help the Distance Education (DE) students succeed in the blended learning approach being provided by the DE educators.

## I. Introduction

Nowadays, instructional technologies are on transition from mobile learning to ubiquitous learning wherein educational materials are accessible anytime, anywhere in any form (text, video and audio) to all educational stakeholders via eLearning platforms. With this innovation, students and enrollment in Distance Education (DE) courses became attractive for most learners. Conversely, the report on course drop out and failure rates is more incessantly increasing in this mode of learning. In this light, the academe need to develop tools and methods that will explore data coming from the eLearning software which can support teachers and students to take action based on the evaluation of educational data.

Policy makers and administrators should include analytics, user modeling, user profiling and clustering, domain modeling, relationship mining and data visualization to unveil outcome-oriented actionable insights from specific learning behaviors [1]. Consequently, some educational institutions (e.g. University of the Philippines Open University, Mindanao State University, California State University, Monash University in Australia) were implementing Learning Management System (LMS) to manage the courses offered in the Internet [3]. In 2015, eLearning industry reported that 74% of the companies currently use LMS and Virtual classroom, webcasting and video broadcasting [4]. Some of the most popular open source LMS includes Edmodo and Modular Object-Oriented Dynamic Learning Environment (Moodle) [8]. Nevertheless, based on the studies, MOODLE was the most recommended LMS because the administration and control can be handled by the institution to do further analytics such as tracking web logs of students for further monitoring of their progress and other activities [2]. According to Romero et. al., the application of data mining in e-learning is not much different than any other application area [9]. However, there are some important issues that make data mining in e-learning different than in the others such as (1) data; (2) objective and (3) techniques. In other web-based systems the data used is normally a simple web server access log, but in e-learning there is much more information available about the student's interaction such as details on each online assessment task (assignment, quiz, discussion forum and chat) [6]. Conversely, Picciano suggested that LMS's should provide constant monitoring of student activity whether there are responses, postings on a discussion board, accesses of reading material, completions of quizzes, or some other assessment. Thus, university should analyze the data gathered from the LMS users which are stored in the web server to discover knowledge that will enhance the students' online experience [7]. ECAR Working Group (2015) suggested that educators can tap

This paper is prepared exclusively for International Conference on Computer, Engineering, Law, Education and Management 2017 [ICCELEM 2017] which is published by ASDF International, Registered in London, United Kingdom under the directions of the Editor-in-Chief Dr. K Kokula Krishna Hari and Editors Dr. Daniel James, Dr. Saikishore Elangovan. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honoured. For all other uses, contact the owner/author(s). Copyright Holder can be reached at copy@asdf.international for distribution.



more dynamic data produced from a range of instructional technologies (such as LMS event log data, electronic gradebook data, attendance data, library data, etc.) for learning analytics which when combined with traditional measures—allow for a more nuanced and personalized analysis [5]. Some of the challenges of implementing data mining and learning analytics, include high cost of collection, storage, development of algorithms, interoperable administrative and learning systems (systems/data types). As such, their report recommends researchers to combine the data types, with acceptable validity, privacy, and ethical standards applied, for improved predictive power [1].

Apparently, in one of the student workshops conducted by the University of Lincoln, the students cited some ideas on capabilities they would like to see in a learning analytics application. They included notifications on grades and progress toward objectives; the ability to give immediate feedback to lecturers and professors in order to improve the course; and reading list functionality that presents metrics on how students engage with the texts [10].

Currently, little research has been conducted that focus on university portal learning analytics that will lead to the prediction of academic performance of the DE students. Thus, this study focused on providing a Decision Support System (DSS) that integrates Multiple Linear Regression (MLR) model of data mining for portal providers and users to analyze and predict the performance of distance students scientifically.

## II. Framework of the Study

Fig.1 illustrates the framework of the study. It consists of four (4) major phases 1) development of a student performance predictive model. 2) testing and implementation of the predictive model, 3) development of the decision support system to identify at risk-students for early intervention and attrition prevention and 4) evaluation of the developed software by the respondents in terms of the following: functionality, usability, reliability and portability of the output.

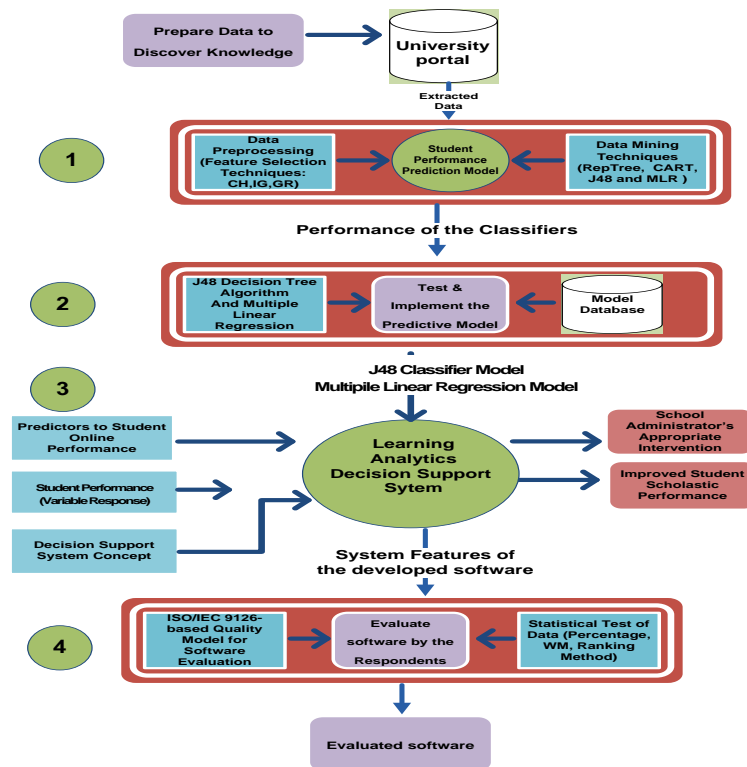


Figure 1. Framework of the Study

During the first phase, the author utilized the data sets from the university Academic Institutions Management Systems (AIMS) and the students' usage history while accessing the Polytechnic University of the Philippines (PUP) eMabini Learning Portal. After which, the author performed steps to preprocess the data then converted the extracted data into the required format by

Waikato Environment for Knowledge Analysis (WEKA) tool. Also, to generate the students' performance prediction model, the author conducted series of experiments to evaluate the appropriate classification for predicting students' final rating based on their usage data in the university portal. During the next phase: testing and implementation of the model, the author identified interesting rules and patterns for decision-making. The author repeated the data mining and the pattern analysis process if she considered that the results were not remarkable. For the third phase of the study, the predictors to Student Online Performance, the data mining techniques and decision support system concept were integrated to develop the software as shown on Fig.1.

The independent variables in this study were the records of student's performance in online assessment tasks posted by the course specialists in the university portal. The expected output was the developed student performance model with MLR. With this model, it may give students considerable time and opportunity for early interventions to improve his scholastic performance and for the Distance Education providers to lessen the drop-out rate. During the fourth phase of the study, the developed software was evaluated by the respondents in terms of functionality, usability, reliability and portability of the output based on the ISO 9126 Quality Model for Software Evaluation. The author used the statistical tools such as Weighted Mean, Ranking Method and Percentage to summarize and analyze the respondents' evaluation on the developed software.

### III. Results and Discussions

#### A. The Multiple Linear Regression Model

The original database for MLR was divided into two using the 80:20 rule - the training and validation dataset which consisted of 19 instances and the test dataset which consisted of 7 instances. Equation 3.0 shows the MLR equation generated using Weka in a 10-fold cross validation and a confidence factor of 0.25 without pre-processing of attributes. This was named as MLR Model A in this study.

MLR Model A in equation form is:

$$\text{Final Rating} = 0.2082 \times \text{Assign}_1 + 0.1987 \times \text{Assign}_2 + 0.205 \times \text{QUIZ}_1 + 0.1963 \times \text{QUIZ}_2 + 0.1986 \times \text{Assign}_3 + -0.4605 \tag{3.0}$$

Fig. 2 described the attributes selection output. It can be gleaned that Log\_Count got the highest influence followed by Mat\_Access\_Count, Activity\_Rating and Final\_Grade.

```

==== Run information ====
Evaluator:  weka.attributeSelection.CfsSubsetEval
Search:    weka.attributeSelection.BestFirst -D 1 -N 5
Relation:  LMS_Prediction_Merged_v6-weka.filters.unsupervised.attribute.Remove-R1-8,10,12,14,16-17,19-20
Instances: 248 ; Attributes: 5
Log_count, Mat_Access_Count, Activity_Ratng, Exam_Points, Final_Grade

Evaluation mode: 10-fold cross-validation
==== Attribute selection 10 fold cross-validation seed: 1 ====
number of folds (%) attribute
10(100 %) 1 Log_Count; 1( 10 %) 2 Mat_Access_Count;
0( 0 %) 3 Activity_Rating; 10(100 %) 5 Final_Grade
    
```

Figure 2. Attribute Selection Output for MLR

From the processing of attributes making use of the same validation and confidence factor, the MLR model generated is shown in Equation 4.0. This was named as MLR Model B in this study.

MLR Model B in equation form is:

$$\text{Exam\_Points} = -1 \times \text{Activity\_Rating} + 2 \times \text{Final\_Grade} + 0 \tag{4.0}$$

Table I shows the error measures of the two MLR models generated with and without preprocessing of attributes. It compares the fitting of the models as to the differences between the observed values and the model's predicted values. The correlation coefficient of MLR Model A is higher than Model B. On the other hand, MAE and RMSE of MLR Model B was greater than MLR Model A's.

Table I Error Measures of the MLR Models

	MLR Model A	MLR Model B
Correlation Coefficient	0.9999	0.9188
Mean Absolute Error (MAE)	0.1046	8.9507
Root Mean Squared Error (RMSE)	0.1383	11.6134

Fig. 3 examines the difference between the performance of the two MLR model as far as evaluation on test set is concerned. It could be gleaned from the figure that MLR Model B had a higher MAE and RMSE than that of the MLR Model A.

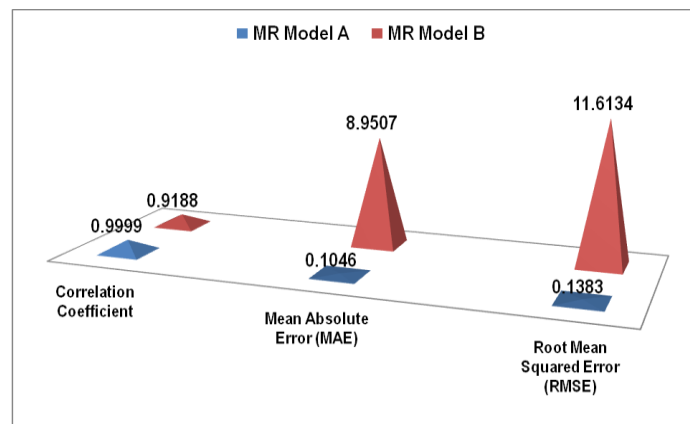


Figure 3. Performance of the MLR Models based from Evaluation on Test Set

## B. Evaluation of the Developed Software by the Respondents

As the developed software will be used for wider DE learners and providers, the author asked select active portal users to evaluate the system in terms of functionality, usability, reliability and portability of the output. After exploring the demonstration, the author encouraged them to fill-out the online survey form.

There were one hundred ninety-one (191) online portal users who filled-out the survey form. Below is the highlight of the information that were gathered from the participants.

Table II described the summary of the evaluation of the developed software as perceived by the respondents. The ratings in this table indicates that the respondents rated the developed software as "Moderately Acceptable" which signifies that it is functional, usable, reliable and portable for the DE stakeholders.

Table ii. Summary of the evaluation of developed software by the respondents

Parameters	Total Weighted Mean	Interpretation
Functionality	4.43	Moderately Acceptable
Usability	4.34	Moderately Acceptable
Reliability	4.29	Moderately Acceptable
Portability	4.32	Moderately Acceptable
Overall Mean	4.34	Moderately Acceptable

In terms of system functionality, the respondents found that the system can shut out access from people who are not part of the course. It provides online submission of assignments where the result can be evaluated by the professor then recorded in the database. This is an indication of an effective integration management that should be maintained by the system. In terms of system usability, the system rated as “Moderately Acceptable” of most of the respondents. This indicates the versatility of the developed software when it comes to usability as it gives opportunity to the users to recapitulate, retrieve and interact with the system whenever the user desires using his available technology gadget. In terms of reliability, the respondents rated all questions as “Moderately Acceptable”. It signifies that the developed system provides confidentiality in each user account and it has more secured delivery and distribution of information to its intended users. These indicate that it generates notifications from authorized users only; provides consistent result and response correctly when encountered failure. In terms of portability it is rated as “Moderately Acceptable”. This indicates that the system can be accessed from one gadget (smart phone, tablet, laptop and personal computer) to another. Thus, it allows the users (OUS Administrator, course specialists and learners) to view course materials and online assessment tasks with one login password in any computer or mobile device which promotes higher system flexibility to the DE learners who were predominantly part-time students but fulltime employees. The ratings on this parameter exhibit the portability of the system which also promotes robustness of the developed model for predicting the online performance of the DE learners.

#### IV. Conclusion

The results demonstrated that the generated MLR model can be harnessed to develop the Learning Analytics Decision Support System which may provide powerful educational tool that can analyze and predict the performance of the learners in the flexible digital learning environment. During the testing and simulation of real institutional data, the developed software displayed the same output with that of the two reliable application programs the Microsoft Excel and WEKA. The respondents rated the developed software as “Moderately Acceptable” with overall mean of 4.31 which signifies that it is functional, usable, reliable and portable for the Distance Education (DE) stakeholders. For future research, the author may concentrate on greater number of instances using the other variables and may explore other data mining algorithms.

#### References

1. Bienkowski, M., Feng, M., & Means, B. 2012. Center for technology in learning sri international, U.S. Department of Education, Office of Educational Technology. Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief. 1-77.
2. Cavus, N. and Zabadi, T. 2014. A comparison of open source learning management systems, Elsevier Ltd. Procedia - Social and Behavioral Sciences 143, 521 – 526.
3. Comendador, B.E.V.and Guillo, A. C., 2014. A community cloud-based course management system using platform as a service (paas) for higher educational institutions, International Journal of Information and Education Technology, (IJJET), ISSN: 2010-3689 International Association of Computer Science and Information Community Press, 4
4. eLearning Industry 2015. Retrieved from Elearning Statistics and Facts For 2015 You Need to Know, on January 13, 2017
5. ECAR Working Group 2015. The predictive learning analytics revolution leveraging learning data for student success, Educase Center for Analysis and Research. Retrieved from <http://net.educause.edu/ir/library/pdf/ewg1510.pdf>
6. Jailia, M. & Tyagi A., 2013. Data mining: a prediction for performance improvement in online learning systems, International Journal of Advanced Research in Computer Science and Software Engineering, 3(7), 628-635.
7. Picciano, A. G., 2014. Big data and learning analytics in blended learning environments: benefits and concerns, International Journal of Artificial Intelligence and Interactive Multimedia, 2(7), 35-43.
8. Red, E. R., Cancino, G. C. C., Hanrath, M. J. W. V., & Ricardo, J. G. E., 2014. Analysis of the offline and online data of students using clustering and classification techniques, International Journal of the Computer, the Internet and Management, 22(3).
9. Romero, C., Ventura, S., & García,E. 2010. Data mining in course management systems: Moodle case study and tutorial, Computers & Education, 51(1), 368-384.
10. Sclater, N., 2015. “What do students want from a learning analytics app?”



ISBN	978-81-933584-7-4
Website	www.iccelem.com
Received	05-June-2017
Article ID	ICCELEM002

VOL	01
eMail	iccelem@asdf.res.in
Accepted	20-July-2017
eAID	ICCELEM.2017.002

# Sentiment Analysis on Service Complaints using Sarcasm and Emoticon Detection

Ria A Sagum<sup>1</sup>, Clowie Fel Y Bingco<sup>2</sup>, J Wesley P Florentino<sup>3</sup>

<sup>1,2,3</sup>Department of Computer Science, College of Computer and Information Sciences,  
Polytechnic University of the Philippines, Sta. Mesa, Manila

**Abstract-** This study developed a sentiment analyzer that utilizes sarcasm detection. A multilingual language model was used to classify complaints with the use of Probabilistic Model with Kneser Ney smoothing to improve the system's accuracy. Emoticons were also detected to include in the sarcasm detection of a service complaint. The researchers made use of dataset from servicing company or government agency's service complaint to ensure that the research would be based on real life data. The system gained precision of 89%, recall of 92 % and F1-Score of 90%. The study shows that the accuracy of the Sarcasm Detection increased upon the integration of Emoticon Detection and resulted to the increase in sentiment analysis of the system.

**Keywords:** Emoticon Detection, Language Model, Sentiment Analysis, Sarcasm Detection, Sarcasm and Emoticon Detection

## I. Introduction

Sentiment analysis of review sites and online forums has been a popular subject for several years in the field of natural language processing. [1] Before the internet awareness became widespread, many of people used to ask their friends or neighbors for opinion of a good electronic products or a food before buying it or going for it. With the growing availability and popularity of opinion-rich resources such as online review websites and personal blogs, new opportunities and challenges arise as people now can, and do, actively use information technologies to seek out and understand the opinions of others. Unfortunately, these opinion rich resources are available in unstructured format. It has encouraged the analysts to develop an intelligent system that can automatically categorize or classify these text documents.

Sarcasm is a form of art that is marked using sarcastic language and is intended to make its victim the buff on contempt of ridicule. In text mining, automatic detection of sarcasm is considered a difficult problem [2] and has been addressed in only a few studies. Sarcasm can be used to transform the polarity of an apparently positive or negative utterance into opposite [3]. It was suggested by the study of (Sagum et.al.), that sarcasm can be used to increase the accuracy of a sentiment analyzer. Detecting sarcasm and emoticon in text is a complex process. To recognize sarcasm, tone recognition must also be considered since people express their feelings with high and low pitches [4]. To recognize emoticon, we must know the following: First, emoticons represent body language, which is nonverbal. Second, there has been a lack of sufficient methods for the analysis of emoticons and need to recognize the pattern or identify rather what is the emoticon that has been used [5].

Considering these, the researchers developed a system that will recognize and able to analyze emoticons and sarcastic statements. The researchers strongly agree that it will help to accurately detect emotions, emoticons that were used were based on the article Smiley Face and Text Emoticon Symbols by Beal [6]. The system will be able to classify the polarity of a complaint whether it is positive, negative or neutral. The sentiment analyzer will be using sarcasm detection and emoticon detection as its feature to improve the accuracy of the analyzer, it will analyze service complaints of customers.

This paper is prepared exclusively for International Conference on Computer, Engineering, Law, Education and Management 2017 [ICCELEM 2017] which is published by ASDF International, Registered in London, United Kingdom under the directions of the Editor-in-Chief Dr. K Kokula Krishna Hari and Editors Dr. Daniel James, Dr. Saikishore Elangovan. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honoured. For all other uses, contact the owner/author(s). Copyright Holder can be reached at copy@asdf.international for distribution.

## II. Related Works

The study of Wang et al. [3], explains that automatically detecting sarcasm in twitter is a challenging task since sarcasm transforms the polarity of an apparently positive or negative utterance into its opposite. Previous work focuses on feature modelling of the single tweet, which limit the performance of the task. The sarcasm detection problem is modelled as a sequential classification task over a tweet and his contextual information.

The first study to addresses the use of emoticons to recognize sarcasm was done by Walther et.al [7]. Their participants were asked to read emails including positive or negative messages, followed by a smiley face :-), a sad face :-(, a wink face ;- ) or no emoticon. Messages were ambiguous as to whether they were intended literally or sarcastically. Furthermore, the most sarcastic condition was a positive verbal message with a wink. However, this message–emoticon combination was not significantly more sarcastic than a positive message with a smile, a sad face, or nothing at all. Therefore, the researchers concluded that winks do not actually connote greater sarcasm than other emoticons.

The study of Derks et al. , [8] ran a similar study to Walther et al., examining the same set of emoticons, but included a neutral message condition (in addition to positive and negative), and the participants in the study were the recipients of the emails. In contrast to Walther et al. , the work of Derks et al. showed that emoticons enhanced the valence of a message. Additionally, and again in contrast to Walther et al. messages with a wink face were rated as significantly more sarcastic than those without an emoticon.

Hogenboom established that in order to exploit emoticons in automated sentiment analysis, the researchers first need to analyze how emoticons are typically related to the sentiment of the text they occur in. Interestingly, this positioning of emoticons suggests that it is typically not a single word. [8]

Lexical feature-based classification as first of the types to detect sarcasm is text properties such as unigram, bigram, n-grams, etc. are classified as lexical features of a text. Authors used these features to identify sarcasm, introduced this concept for the first time and they observed that lexical features play a vital role in detecting irony and sarcasm in text. [9] Riloff, et. al. [10] used a well-constructed lexicon-based approach to detect sarcasm and for lexicon generation they used unigram, bigram and trigram features. Barbieri et. al. considered seven lexical features to detect sarcasm through its inner structure such as the intensity of the terms. [11]

Pragmatic feature-based classification as the second type to detect sarcasm uses symbolic and figurative text in tweets is frequent due to the limitations in message length of a tweet. These symbolic and figurative texts are called pragmatic features (such as smiles, emoticons, replies, @user, etc.). It is one of the powerful features to identify sarcasm in tweets as several authors have used this feature in their work to detect sarcasm. Pragmatic features are one of the key features used by Kreuz & Caucci [9] to detect sarcasm in text. The study of Carvalho et. al. [12] used pragmatic features like emoticons and special punctuations to detect irony from newspaper text data.

After thorough understanding on the different studies on sentiment analysis it was then highlighted that sarcasm can help with the accuracy of sentiment analysis, likewise the emoticons can give impact in the accuracy of sarcasm detection. The researchers made use of these features to look on to the accuracy of a sentiment analyzer once sarcasm and emoticon detection were taken into consideration.

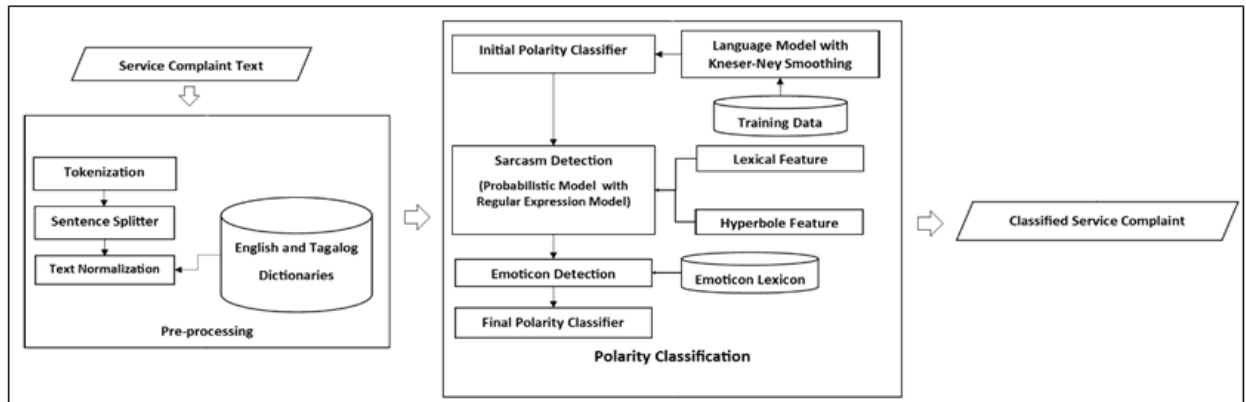
## III. Discussion of the System's Design

The process of the system starts with pre-processing activities (see Figure 1). These tasks include Tokenization, Sentence Splitter and Text Normalization. The researchers used Python's NLTK for this module.

The next phase is the Polarity Classification Module, this module consists of the ff:

- a. **Initial Polarity Classifier**, it distinguishes the polarity of the text whether it is positive, negative or neutral by using the classification of bag-of-words and Language Model. It sets out the polarity for each token in the sentences.

Figure 1: System Architecture



- b. **Language Model**, assigns a probability to a set of string based on its occurrence in text prior processed. The model used N-Grams and a Kneser–Ney smoothing algorithm to improve the probabilities of each gram. The Training Data is a collection of classified complaint sentences based in polarity such as, positive, negative or neutral. The model used up to tri-gram. The reason is that the data that have been used are composed of Tagalog, English and Taglish complaints and the structure of its sentences can be processed correctly using this gram.
- c. **Sarcasm Detection**, this will be implemented after setting the polarity in each text. The algorithm implemented in this process was Probabilistic Model and Regular Expression Model. Using the model, the probability of the word of being negative is from -1.0-0.4, for neutral -0.399-0.399, and for positive it was 0.4-1.0.

e.g. ['Napakagaling', 0.642, Pos], ['ninyo', -0.132, Neu], ['gumawa', 0.539, Pos], ['ng', 0.539, Neu], ['daan', -0.205, Neu], ['.', -0.132, Neu], ['Lubak', -0.742, Neg], ['Lubak', -0.742, Neg], ['parin', 0.143, Neu], [':D', 0.053, Neu]

c.1. **Lexical Feature Classifier**, will be the first to detect sarcasm in text properties using unigram, bigram, trigram, n-grams, etc. each of which are classified as lexical features of a text.

c.2. **Hyperbole Feature**, composed of Intensifier will search for keywords that denote intensity or degree of a given text. These keywords will be used to increase/decrease the positivity or negativity of a certain word/s. The Regular Expression Model was used by the researcher to detect sarcasm by observing patterns that denote sarcastic complaints from a given text. Interjection such as: "wow", "aha", "yay" etc. has a higher chance of being sarcastic, and series of punctuation marks will be used as additional cues for sarcasm detection.

- d. **Emoticon Detection**, will add a value to its polarity if an emoticon placed on a given text. It will add a positive, negative or neutral weight on its context. The emoticon lexicon contains all the emoticons included in the American English twitter corpus study that will be used in this research. In the example, the emoticon polarity's detected shifted to a positive polarity.

e.g. [':D', Neu] -> [':D', Pos]

- e. **Final Polarity Classifier**, computes for the overall polarity weights in the Sarcasm and Emoticon Detection. Adding to it, a rule based was implemented to check if a sentence's sentiment was Positive and it was followed by a negative emoticon and if a sentence's sentiment was Negative and was followed by a positive emoticon, it will have a corresponding result of being sarcastic. Lastly, it will output the classified service complaint whether it is sarcastic or not sarcastic and if an emoticon existed or not. In the example, the first sentence which resulted to a low probability of being positive and the second sentence scoring a higher probability of being negative, resulting to shift the total context of the sentence into a negative sentiment. Now the negative sentence is followed by a positive emoticon. Sarcasm is detected by applying the rule of negative sentence followed by a positive emoticon.

Sarcasm Detection

Napakagaling nyo gumawa ng daan. Lubak lubak parin = NEG

Emoticon Detection

:D = POS

NEGATIVE (Sentiment) + POSITIVE (Emoticon) = Sarcastic

#### IV. Presentation of Results

The performance of the system was measured using the F measure or F1 score (Eq. 1).

$$F1\ Score = 2 * \frac{precision * recall}{precision + recall} \quad (1)$$

F1 score refers to the harmonic mean of precision (Eq. 2) and recall (Eq. 3).

$$Precision = \frac{True\ Positives}{(True\ Positives + False\ Positives)} \quad (2)$$

$$Recall = \frac{True\ Positives}{(True\ Positives + False\ Negatives)} \quad (3)$$

The researchers made use of the table of Dan Jurafsky of Stanford University to classify the output of the stream. It uses True Positive (TP), the number of sentiments classified positively correct by both system and human, True Negative (TN), the number of sentiments that classified positively incorrect by both system and human, False Positive (FP), the number of sentiments classified correctly by system but incorrect by human, and False Negative (FN), the number of sentiments classified correctly by human but incorrect or neutral by system.

Table 1 shows the result of the evaluation of the sentiment in every criterion. The precision of Complaints with Emoticon gained 100%, this is because the dataset that was fed in the system were correctly analyzed by the system. The system recorded 67% for recall due to a complaint that was not recognized by the system and 80% for F1-Score. For Complaints with Sarcasm and Emoticon, it gained 100% high in recall, due to 258 complaints that were correctly recognized, 100% in precision due to one complaint that was not recognize and 100% in F1-Score due to the computed precision and recall. The result of the system's evaluation under the criteria of Plain Complaints for precision, recall and F1-score were 84%, 80% and 82% respectively. And lastly for the Complaints with Sarcasms gained 34%, 42% and 38% respectively for precision, recall and F1 Score. This is due to unclean dataset that was fed in the system during evaluation phase, also the data set lacks complaints that includes sarcastic features.

Table 1: Summary of Results according to Criteria

Criteria	TP	TN	FP	FN	PRECISION	RECALL	FISCORE
Complaints with Sarcasm	38	552	73	52	34%	42%	38%
Complaints with Emoticon	2	712	0	1	100%	67%	80%
Complaints with Sarcasm and Emoticon	258	456	1	0	100%	100%	100%
Plain Complaints	290	298	54	73	84%	80%	82%

Table 2: Overall Systems Performance in Recognizing Sentiments

PRECISION	89%
RECALL	92%
F1 SCORE	90%

By using the sarcasm detection with integration of emoticon detection, in the developed system, the results shows in Table 2 was the Overall Performance of the system for sentiment analysis regardless of the criteria. Based on the study of Ebola etal [13], that generated a rating system for the parameters: precision, recall and F1 score, the system's accuracy rate is "Very Good".



## V. Conclusion

In this study, language model was utilized to determine the sentiment of service complaints. Sarcasm detection was employed through pattern extraction, to improve the overall accuracy of the system. Lastly, the researchers included emoticon detection to gain much higher accuracy in detecting sarcasm in service complaints.

The system was able to serve its purpose: to detect sarcasm and integrate it to the system to accurately analyze a sentiment. If it will be compared to the previous study (Sagum, De Vera, Lansang, Narciso, & Respeto, 2015) Ref [12], where smoothing algorithm was used for the Language Model, the system's precision and recall were quite higher than the previous one. However, there are numerous words that the program was not able to recognize correctly. Human raters typically agree 70% is a Pass rating when it comes to sentiment analyzers. [14] [13] [4] Thus, a sentiment analyzer that has 82% accuracy rate is quite doing well as humans do in analyzing sentiments but still has a lot to learn for improvement. The results of the study showed that the system was able to correctly determine the detection in service complaints. It was rated "Very Good" in performing sentiment analysis.

The results of the research furthermore showed a marginally dense F1-Score. F1-Score ranged from 50-100%, which is considerably satisfactory to excellent rate, but can be improved through a larger set of implementation data. The researchers observed that the F1-Score and Accuracy of the criteria Complaints with Sarcasm and Emoticon is higher than the criteria of Complaints with Sarcasm. The researchers achieved their goal to integrate the emoticon detection to sarcasm detection for better recognition of sarcasm and improved the previous study of Sagum et. al. [4] Researchers found out that the emoticon detection is effective in achieving higher accuracy when integrated in sarcasm detection.

## VI. Recommendations

The implementation set was a set of mixed complaints with different sentiments so the results were not that accurate. The researchers recommend testing the system using a different set of complaints for each sentiment: Positive, Negative and Neutral for much more accurate results of the capability of the sentiment analyzer to classify sarcastic complaints.

The researchers recommend using Chen and Goodman's Modified Kneser-Ney. This smoothing method is tested to be the best smoothing method for these kinds of problems. Probabilistic Model with Modified Kneser-Ney Smoothing needs a much large of training data.

It is also recommended for the future researcher to use other combination of features in sarcasm detection such as Slang detection, Irony detection and Rant detection to improve the degree of accuracy of sarcasm detection in a hybrid manner, thereby increasing the accuracy of sentiment analyzer.

## VII. References

1. Kawathekar, S. A., & Kshirsagar, M. M. (2012, January). Sentiments analysis using Hybrid Approach involving Rule-Based & Support Vector Machines methods. *IOSR Journal of Engineering (IOSRJEN)*, 2(1), 55-58.
2. Nigam, K., & Hurst, M. (2006). Towards a Robust Metric Polarity. In J. G. Shanahan, Y. Qu, & J. Wiebe (Eds.), *Computing Attitude and Affect in*. Springer Netherlands. doi:10.1007/1-4020-4102-0\_20
3. Wang, Z., Wu, Z., Wang, R., & Ren, Y. (2015). Twitter Sarcasm Detection Exploiting a Context-Based Model. *Web Information Systems Engineering (WISE 2015)*, 9418, 77-91.
4. Sagum, R. A., De Vera, J. G., Lansang, P. J., Narciso, D. S., & Respeto, J. K. (2015, March). Application of Language Modelling in Sentiment Analysis for Faculty Comment Evaluation. *Proceedings of the International MultiConference of Engineers and Computer Scientists, IMECS 2015*, 1.
5. Ptaszynski, M., Maciejewski, J., Dybala, P., Rzepka, R., Araki, K., & Momouchi, Y. (2012). Science of Emoticons: Research Framework and State of the Art in Analysis of Kaomoji-Type Emoticons. In U. S. Tiwary, & T. J. Siddiqui (Eds.), *Speech, Image, and Language Processing for Human Computer Interaction* (pp. 234-257). Allahabad, India: IGI, USA.
6. Beal, V. (2016). Smiley Face and Text Emoticon Symbols. Retrieved September 11, 2016, from [www.webopedia.com](http://www.webopedia.com): [http://www.webopedia.com/quick\\_ref/textmessageabbreviations\\_02.asp](http://www.webopedia.com/quick_ref/textmessageabbreviations_02.asp)
7. Walther, J. B., & D'Addario, K. P. (2001). The Impacts of Emoticons on Message Interpretation in Computer-Mediated Communication. *Social Science Computer Review*, 19, 324-347.

8. Derks, D., Bos, A. E., & Von Grumbkow, J. (2008). Emoticons and Online Message Interpretation. *Social Science Computer Review*, 26, 379–388.
9. Hogenboom, A., Bal, D., & Frasinca, F. (2013, March). Exploiting Emoticons in Sentiment Analysis. *SAC*.
10. Kreuz, R., & Caucci, G. (2007). Lexical influences on the perception of sarcasm. (pp. 1–4). *ACL: Proceedings of the Workshop on Computational Approaches to Figurative Language*.
11. Barbieri, F., Saggion, H., & Ronzano, F. (2014). Modelling sarcasm in twitter a novel approach. (pp. 50–58). *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*.
12. Carvalho, P., Sarmiento, L., Silva, M., & De Oliveira, E. (2009). Clues for detecting irony in user-generated contents: oh...!! it's so easy;-). (pp. 53–56). *ACM: Proceedings of the 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion*.
13. Eboña, K. M., O, S., Perez, G. P., Roldan, J. M., Sagum, R. A., & Domingo, I. V. (2013). Named Entity Recognizer (NER) for Filipino Novel Excerpts using Maximum Entropy Approach. Undergraduate Thesis, Polytechnic University of the Philippines, Department of Computer Science, Manila City.
14. Rafael, V. L. (2008). Taglish, or the phantom power of the lingua franca. In *Philippine English*. Riloff, E., Qadir, A., Surve, P., De Silva, L., Gilbert, N., & Huang, R. (2013). Sarcasm as contrast between a positive sentiment and negative situation. (pp. 704–714). *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.



ISBN	978-81-933584-7-4
Website	www.iccelem.com
Received	03-June-2017
Article ID	ICCELEM003

VOL	01
eMail	iccelem@asdf.res.in
Accepted	05-Aug-2017
eAID	ICCELEM.2017.003

# Diagnostic App for Cats and Dogs Diseases using Neuro – Fuzzy Algorithm

Carlo G Inovero<sup>1</sup>, Emmanuel Vincent Gratila<sup>2</sup>, Jane Maelyne C Lopez<sup>3</sup>

<sup>1,2,3</sup>Department of Computer Science, Polytechnic University of the Philippines,  
 A. Mabini Campus, Anonas ST, Sta Mesa, Manila, Philippines

**ABSTRACT:** Dog and cat is considered as beloved pets of most people. But these animals are also prone to different diseases such as colds, ticks and fleas, worms, and fungal infections. Early detection leads to early prevention and cure. Detecting diseases at early stage will enable to overcome and treat them appropriately. Identifying the treatment accurately depends on the method that is used in diagnosing the diseases. This study entitled “Diagnostic App for Cats and Dogs Diseases using Neuro – Fuzzy Algorithm” a developed mobile-based application could recognized dogs and cats diseases using neuro-fuzzy algorithm. It aimed to test the accuracy performance of the neuro-fuzzy algorithm on the mobile app. The researchers use Android Studio as the coding platform and Java as the programming language. The developed mobile app run on android version KitKat or better versions. The researchers used experimental method of research which aimed to evaluate the accuracy performance of the mobile with neuro-fuzzy algorithm in diagnosing dog’s diseases, cat’s diseases in terms of precision, recall, and f-measure. The accuracy performance rate of the system was measured through series of experimentation and with the help of our expert. The researchers used 171 for dogs’ diseases and 124 diseases for cats to test the performance accuracy of the said mobile app with neuro-fuzzy algorithm. The study attained the overall accuracy performance rate of the mobile app with neuro-fuzzy with of 87% in diagnosing dogs’ diseases and 90% in diagnosing cats’ diseases. The overall accuracy performance is 88.50%. Hence, the researchers concluded that the Diagnostic App for Cats and Dogs Diseases using Neuro – Fuzzy Algorithm is very high. The developed mobile App can diagnosed dogs’ and cats’ diseases and could advise what proper treatment could be done for every illnesses. It could be a guide for the pet’s owner in taking good care of their loving animal. We, therefore recommend to use the said mobile app in diagnosing dogs’ and cats’ diseases.

**Keywords:** Neuro Fuzzy, Diagnostic Mobile App, Pet App

## 1. Introduction

According to Veterinary Pet Insurance, there are 10 common diseases that affects dogs and cats. Common diseases in dogs and cats are skin allergies, ear infection, non-cancerous skin mass, skin infection, arthritis, vomiting/upset stomach, periodontitis/dental disease, diarrhea/intestinal upset, bladder or urinary tract infection, soft tissue trauma (bruise or contusion), excessive thyroid hormone, upper respiratory infection, and lymphoma. (Association, 2016). Preventive healthcare involves a multi-faceted approach that includes veterinary evaluation of your pet's overall health and risks of disease or other health problems.

Most of the pet owners may turn to internet and try to diagnose and treat their pet’s conditions themselves (Zander, 2016). However, different sites provide thousands of different advices stored online and finding the right one can be difficult especially if it is coming from different sources. Applications on market like EasyVet is made especially for Veterinarians (Technologies, 2015).

This paper is prepared exclusively for International Conference on Computer, Engineering, Law, Education and Management 2017 [ICCELEM 2017] which is published by ASDF International, Registered in London, United Kingdom under the directions of the Editor-in-Chief Dr. K Kokula Krishna Hari and Editors Dr. Daniel James, Dr. Saikishore Elangovan. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honoured. For all other uses, contact the owner/author(s). Copyright Holder can be reached at copy@asdf.international for distribution.

Diagnosing the diseases and identifying the right treatment can remedy the pets' illnesses. Detecting at early stage of pets diseases prevented serious illness and could be treated properly (Patra, Sahu, & Mandal, 2010). A Diagnosis Expert System (DExS) could be served as a guide in identifying diseases and suggests methods for curing diseases.

According to the study of (Andrews, et al., 2015), veterinarians believe there is a strong desire for mobile technology in veterinary medicine and the use of this technology will allow them to practice more effectively. Results showed mobile devices are prevalent and widespread among veterinarians with more than sixty percent surveyed strongly agreed mobile technology will advance patient care, client communication, and improve access to clinical data and medical literature.

## 2. Background and Its Problem

A Care for Animal Organization believes that animals are wonderful part of people's lives, they bring joy, happiness, and give unconditional love. They bring sorrow when they leave, but most of all, they leave with cherished memories of true friendship." (CFA, 2017)

Pet health care is one thing pet owners want to get right (Haight, 2013). Regular veterinarian visits are not only good for your pet, but can be good for the pet owner's wallet as well. Early detection of illness, like food allergies and urinary tract infections can help prevent or cure these problems, before they become serious or extremely expensive.

Current research like Virtua-Vet done by (Floresca, Jaymalin, Taguba, & Zapanta, 2010) is a big help in preventing diseases on cats and dogs, this research uses Natural Language Understanding Techniques in determining keywords on user input questions. However, data available are very limited since it only focused in determining common diseases. This research is very useful to the pet owners who seek expert advice for their pet but unable to afford the potential costs for veterinary consultation.

Almost same study has been found but it focus on the diagnosis of diabetes. In the study of (Morales & Tomines, 2016), they developed a mobile application that runs on an android operating system. It calculates a user's risk in having diabetes through a set of risk factors and symptoms. This study aims to provide self-awareness and early detection of diabetes to avoid further complications. This study uses fuzzy logic algorithm and genetic algorithm.

This research aimed the following objectives:

1. Developed a mobile application that can diagnosed dogs and cat's diseases.
2. Implemented the Neuro-Fuzzy Algorithm in developing the mobile application.
3. Tested the accuracy performance of the Neuro-Fuzzy Algorithm in the developed mobile application to diagnosed dogs and cat's diseases in terms of Precision, Recall and F-Measure.
4. Attained the overall accuracy performance of the Neuro-Fuzzy Algorithm in the developed mobile application to diagnosed dogs and cats diseases

This study will be a great help to pet owners and veterinarians. Pet owners specifically those who have dogs and cats as their home pets will be the one who will highly benefited in this system. It can help them to save money because they don't need to visit a veterinary clinic in case their pets are sick. For veterinarians, this developed system will help them to serve as a tool in diagnosing pet's diseases. It also makes their work easier when giving a diagnosis about pet disease.

Future Researchers who want to make a study about an android based expert system in diagnosing a pet disease can use this study as their basis. This study can also be improved by other researchers for better development of the system. In the field of Computer Science, this study is beneficial because due to the continuous trend of technology today, having an expert system in diagnosing different pets' diseases will be a great contribution.

## 3. Methodology

### 3.1. Research Methodology

This study used experimental research design. The experimental research approach is a collection of research designs which use manipulation and controlled testing to understand causal processes. Generally, one or more variables are manipulated to determine their effect on a dependent variable. This is an experiment where the researcher manipulates one variable, and control/randomizes the rest of the variables. It has a control group, the subjects have been randomly assigned between the

groups, and the researcher only tests one effect at a time. It is also important to know what variable(s) you want to test and to measure.

Experiments are conducted to be able to predict phenomenon. Typically, an experiment is constructed to be able to explain some kind of causation. Experimental research is important to society - it helps us to improve our everyday lives. (Blakstad, 2008). The researcher to maintain control over all factors that may affect the result of an experiment. In doing this, the researcher attempts to determine or predict what may occur.

### 3.2. System Architecture

In the system architecture that was presented in Figure 6, the first part is the user must select the type of pet, choose from list of symptoms and answers the question. All the selected symptoms and the answers in the Q & A will then go to the neuro-fuzzy model. The output of the developed system will be the recognized name of disease, its description and suggestion.

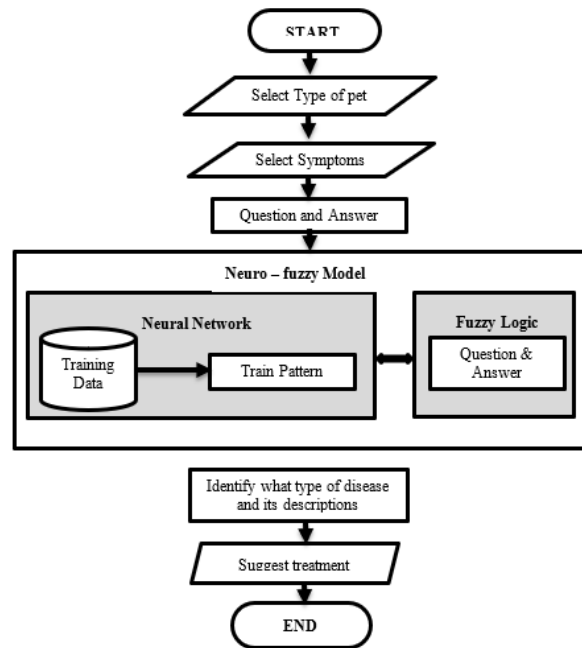


Figure 2 – System Architecture

### 4. Results and Discussions

Name of Disease	True Value	TP	FP	FN	Precision	Recall	F - Measure
Canine Parvovirus	18	15	5	3	75%	83%	79%
Kennel Cough	17	15	2	2	88%	88%	88%
Distemper	17	15	0	2	100%	88%	94%
Demodectic Mange	17	16	4	1	80%	94%	86%
Sarcoptic Mange	17	15	3	2	83%	88%	86%
Leptospirosis	17	14	6	3	70%	82%	76%
Ehrlichiosis	17	13	1	4	93%	76%	84%
Pyoderma	17	14	0	3	100%	82%	90%
Ear Mites	17	17	0	0	100%	100%	100%
Seborrhea	17	15	1	2	94%	88%	91%
<b>AVERAGE</b>					88%	87%	87%

Table 4.1 above is the depicted the performance of the mobile app in diagnosing common dog's illnesses in terms of precision, recall and F-Measures. Earmites got the rate of 100% in terms of Precision, Recall and F-measure. Diseases Distemper and Pyoderma got the 100% rating in terms of Precision, while the second highest in terms of recall is Demodectic Mange disease. Distemper got the second highest score in F measure with a rate of 94%. Leptospirosis got the lowest rate of 70% and 76% in precision and F-Measure respectively. The Ehrlichiosis disease obtained the low rate of 76% in terms of Recall. The average for precision recall and measure of the mobile app for diagnosing dog's diseases is 88%, 87% and 87% respectively.

<b>Name of Disease</b>	<b>True Value</b>	<b>TP</b>	<b>FP</b>	<b>FN</b>	<b>Precision</b>	<b>Recall</b>	<b>F - Measure</b>
Urinary Tract Infection	21	19	5	2	79%	90%	84%
Poisoning	21	16	4	5	80%	76%	78%
Feline Worms	21	19	0	2	100%	90%	95%
Fungal Dermatitis	21	21	1	0	95%	100%	98%
Ear Mites	20	19	0	1	100%	95%	97%
Respiratory Infection	20	18	2	2	90%	90%	90%
<b>AVERAGE</b>					<b>91%</b>	<b>90%</b>	<b>90%</b>

The performance of the mobile app in diagnosing common cat's illnesses in terms of precision, recall and F-Measures showed in table 4.2. Feline worms and Ear mites with the rate of 100% are precisely identified by the mobile app while Fungal Dermatitis got 100% rate in Recall. The highest F-Measure is Earmites with the rate of 97%. The performance of the mobile app in diagnosing cat's diseases is 91%, 90% and 90% for precision, recall and F Measure respectively.

## 5. Conclusions and Recommendations

Table 5.1 shows the overall accuracy performance of the Mobile App with the Neuro-Fuzzy Logic Algorithm in diagnosing the dogs' and cats' diseases. Having an accuracy of 87% in diagnosing dogs' diseases and an accuracy of 90% in diagnosing cats' diseases, the researchers attained an overall accuracy of 88.50%. Also, the Neuro-Fuzzy Logic Algorithms performs well in determining the pets' diseases. Hence, researchers concluded that the accuracy of the system in diagnosing a disease is very high [Artigo et al. 2015].

<b>PET</b>	<b>Percentage</b>
<b>Dog's Diseases</b>	87.00%
<b>Cat's Diseases</b>	90.00%
<b>Average</b>	<b>88.50%</b>

The developed mobile App with Neuro-Fuzzy Algorithm can diagnosed dogs' and cats' diseases and could advise what proper treatment could be done for every illnesses. It could be a guide for the pet's owner in taking good care of their loving animal. We therefore recommend to use the said mobile app in diagnosing dogs' and cats' diseases

Lastly the researchers also recommend the following:

1. Provide if necessary a specific/unique symptom for every disease to correctly recognize a disease.
2. Put an algorithm that will automate the adding of rules based on the patterns provided.
3. Make the symptoms of every disease more specific and should be understand by the user.
4. For future studies, researchers may improve the system and test the significant difference of the experts' diagnosis and systems' diagnosis.

## References

1. Abiyev, R. H., & Abizade, S. (2015, December). Diagnosing Parkinson's Diseases Using Fuzzy Neural System. *Computational and Mathematical Methods in Medicine*.
2. Andrews, C., Bulloch, L., Dennison, T., Elder, J., Mitchell, A., Rivenbank, M. T., . . . Gallicchio, V. S. (2015). *Mobile Technology in Veterinary Clinical Medicine*. Department of Biological Sciences, College of Agriculture, Forestry & Life Sciences, Clemson University, USA.
3. Artigo, P., Faller, A., Macadangang, J., & Martizano, A. (2015). Dietfix: a dietary fitness coach and planner implementing set partitioning in text aggregation and reinforcement learning for adaptive text generation. ., Polytechnic University of the Philippines.
4. Association. (2016). American Veterinary Medical. Retrieved from Preventive Pet Healthcare: <https://www.avma.org/public/PetCare/Pages/Preventive-Pet-Healthcare.aspx>
5. Blakstad, O. (2008). Explorable. Retrieved from Experimental Research: <https://explorable.com/experimental-research>
6. Carse, S. J. (2013, November). A Veterinary Diagnosis Expert System for Remote Use. Bachelor of Science of Rhodes University.
7. CFA. (2017). Care for Animals. Retrieved from <http://www.careforanimals-nc.org/>
8. Explorable. (2017). Explorable.com. Retrieved from Stratified Sampling Method: <https://explorable.com/stratified-sampling>
9. Floresca, L., Jaymalin, C., Taguba, Q., & Zapanta, K. J. (2010). Dog Illness Diagnostic System for Dog Owners and Veterinarians: An Assessment.
10. Fu, Z., Xu, F., Yun, Z. ., & Zhang, X. S. (2005). Pig-Vet: A Web-Based Expert System for Pig Disease Diagnosis. *Expert Systems with Applications*, 29, 93-103.
11. Haight, M. (2013). Pet Health Care: What To Do When You Can't Afford Vet Care. Retrieved from Dancing Dog Blog: <http://www.dancingdogblog.com/2013/04/pet-health-care-what-to-do-when-you-cant-afford-vet-care>
12. Investopedia. (2017). Retrieved from [http://www.investopedia.com/terms/stratified\\_random\\_sampling.asp](http://www.investopedia.com/terms/stratified_random_sampling.asp)
13. Morales, E. M., & Tomines, A. D. (2016). Diabetest: A Mobile Based Fuzzy Expert System for Diagnosis of Diabetes. Undergraduate Thesis, Polytechnic University of the Philippines, Manila.
14. Munirah, Y., Suriawati, S., & Teresa, P. (2016). Design and Development of Online Dog Diseases Diagnosing System. *International Journal of Information and Education Technology*.
15. Patra, P. K., Sahu, D. P., & Mandal, I. (2010). . An Expert System for Diagnosis Of Human Diseases. *International Journal of Computer Applications*.
16. Publishing, C. (n.d.). what-when-how. Retrieved from Adaptive Neuro-Fuzzy Systems (Artificial Intelligence): <http://what-when-how.com/artificial-intelligence/adaptive-neuro-fuzzy-systems-artificial-intelligence/>
17. Statistics. (2017). Statistics How To. Retrieved from Slovin's Formula: What is it and When do I use it?: <http://www.statisticshowto.com/how-to-use-slovins-formula/>
18. Technologies, A. (2015). Easyvet Veterinary Drug Index. Retrieved from Google Play: <https://play.google.com/store/apps/details?id=com.aitrich.Easyvet&hl=en>
19. USC. (2017). Organizing Your Social Sciences Research Paper: Independent and Dependent Variables. Retrieved from University of Souther California Libraries: <http://libguides.usc.edu/writingguide/variables>
20. Zander, J. A. (2016). Top Challenges for Veterinary Practices.



ISBN	978-81-933584-7-4
Website	www.iccelem.com
Received	05-June-2017
Article ID	ICCELEM003

VOL	01
eMail	iccelem@asdf.res.in
Accepted	03-Aug-2017
eAID	ICCELEM.2017.003

## Automatic Classification of Consumer Queries using LSTM

Long Cai<sup>1</sup>, Kokula Krishna Hari K<sup>2</sup>, Rajkumar Sugumaran<sup>3</sup>

<sup>1</sup>Research Scholar, Association of Scientists, Developers and Faculties, Hong Kong

<sup>2</sup>Secretary General, ASDF International, London, United Kingdom

<sup>3</sup>Vice President, Human Resource Management, Techno Forum Group, Bangkok, Kingdom of Thailand

**Abstract:** In this digital era, it is very important to understand the consumer needs while dealing with large volume of data. In this paper, we focus on consumer queries and complaints. It is a difficult task to manually sit and arrange the queries and complaints in forums or online discussion sites related to a specific topic. We propose a method which automatically classifies the queries posted by consumer to its correct class. The system does this classification by using a technique called Long Short-Term Memory(LSTM). The LSTM network has the capability of learning long-term dependency features directly from the dataset without any manual effort. The model showed considerable accuracy when tested with validation data.

**Keywords:** Artificial Intelligence; Deep Learning; Feed Forward Network; Recurrent Neural Network; Long Short-Term Memory.

### INTRODUCTION

Machine learning algorithms need pre-defined features to work. It is a very difficult task to identify salient features from data. Domain data knowledge is very essential for applied machine learning. The process of transforming data into features is called feature engineering. The process of feature engineering is a time-consuming method. In deep learning, the neural network will learn the features automatically from raw data.

Feed Forward Network (FNN) is a type of Artificial Neural Network (ANN) where the information goes in forward direction only. The simple FNN has no hidden layers. In case of FNN with one perceptron, the computed output will be the sum of the product of their weights. When we step back and look at the data, we will understand the pattern of that data. By storing these patterns, we can predict the next sequence by just seeing the previous sequence.

Recurrent Neural Network (RNN) stores information in the memory over time. The vanishing gradient problem in RNN makes it difficult to store long term dependencies. The network is trained using backpropagation algorithm. It uses chain rule that gives derivatives or partial derivatives of a function.

RNN requires complex architecture than non-recurrent networks. The chain rule requires lots of computation. The output of RNN is not only used for computing recurrent value but also for computing next value for time periods. In deep neural networks, there are lot of hidden layers. The fundamental flaw of recurrent neural network is the number of multiplications required to compute the updated weights. The computed coefficients or weights in the past hidden layers are small numbers. So, it's hard for RNN to learn from the past. Consider this sample sequence A saw B, B saw C, C saw D. In this example, we need to

This paper is prepared exclusively for International Conference on Computer, Engineering, Law, Education and Management 2017 [ICCELEM 2017] which is published by ASDF International, Registered in London, United Kingdom under the directions of the Editor-in-Chief Dr. K Kokula Krishna Hari and Editors Dr. Daniel James, Dr. Saikishore Elangovan. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honoured. For all other uses, contact the owner/author(s). Copyright Holder can be reached at copy@asdf.international for distribution.



predict the next sequence after 'A'. 'A' will strongly vote for 'saw' and 'B' will vote for comma. The word 'saw' has equal chances of predicting B, C or D. So, there are chances to make wrong prediction. To predict correctly, we need to see what happened in the previous steps.

LSTM is a type of RNN with a set of gates to control the flow of information. The gates will select and forget information when it enters the memory. The on/off gates will decide what to release as prediction and what to keep internal.

The dataset used for this work is US Consumer Finance Complaints. It is about issues people experienced in marketplace. The 'issue' and 'sub issue' columns in the data shows the problems faced by consumers. The product column shows products like mortgages, student loans, payday loans, debt collection, credit reports, and other financial products and services. Each record or sequence is a combination of 'issue' and 'sub issue' column. The product column corresponding to each record is taken as class label.

## RELATED WORK

Oguzhan Gencoglu [1] proposed a method to categorize the messages from Finland's largest online health forum. It is to reduce the manual effort in managing messages in the forum. He used a Naïve Bayes classifier to classify messages into 16 categories.

The Search result diversification enables the modern-day search engines to construct a result list that consists of documents that are relevant to the user query and at the same time, diverse enough to meet the expectations of a diverse user population. However, all the queries received by a search engine may not benefit from diversification. Sumit bhatia, Cliff Brunk and Prasenjit Mitra [2] proposed an idea to analyze web search queries and classify those queries into one of the classes. They achieved Strong classification results for this classifier.

Lorenzo A Rossi and Omprakash Gnawali [3] analyzed the discussion threads from coursera forums. They investigated several language independent features to classify the discussion threads based on the types of the interactions among the users. The features related to structure, popularity, temporal dynamics of threads are extracted.

Bernard J. Jansen and Danielle Booth [4] proposed a methodology to classify automatically Web queries by topic and user intent. This technique can be used for real time query classification of web searches.

D. Irazú Hernández, Jansen Parth Gupta, Paolo Rosso and Martha Rochagy [5] proposed a method to automatically extract features from corpora and analyzed the distribution of features and used Naïve Bayes and SVM to classify them.

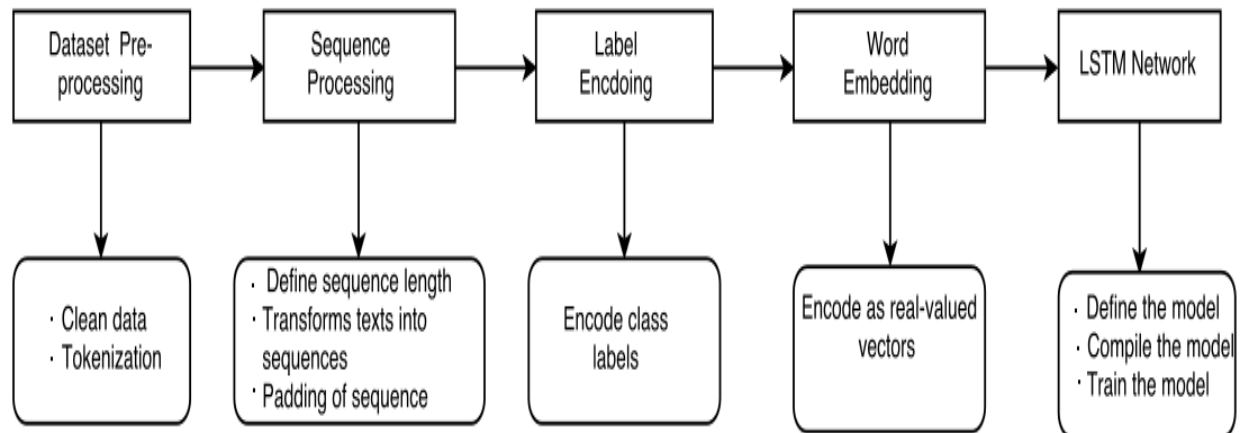


Fig. 1. Workflow of Model

Kristof Coussement, Dirk Van den Poel [6] introduced a technique to improve complaint-handling strategies through an automatic email-classification system that distinguishes complaints from non-complaints. This methodology reveals linguistic style differences between complaint emails and others.

## SOLUTION APPROACH

The workflow of the model is shown in figure 1. The approach of sequence classification model is explained below.

**Dataset Pre-Processing:** Tokenization and stop word removal are the common pre-processing steps. Each record is converted into unit-gram tokens. Keras Tokenizer API is used for tokenization and other basic filtering of text. The most common words are removed from the raw text. Other unwanted symbols and numbers are also removed from the data using regex operations.

**Sequence Processing:** The first step is to transform each record to sequences. A vocabulary is created based on tokens. Each word in the dictionary is represented with a unique number. The next step is to pad the sequence length to the defined size. If the sequence length is smaller than the defined size, zeros will be added to pad the sequence. We can discard it if the size is higher than maximum sequence length.

**Label Encoding:** The algorithm will not be able to read class labels. The class labels are transformed into an array of numbers.

**Word Embedding:** The process of representing words in a continuous vector space based on position of words. This representation gives semantic similarity between words. The distributed representation of words are given as an input to the embedding layer.

**LSTM Network:** The model is defined by giving the number of memory neurons, activation function etc. We used SoftMax as the activation function. The total dimension represents the features. These features are converted into memory units. It is a fully connected network. LSTM network learns what to select and forget from the features. The model is then compiled by defining the optimization algorithm and loss function. Adam optimizer algorithm is used in the network. It is then fitted to the model. The model is evaluated with the validation data.

Overfitting is the main problem in LSTM networks. The network will not be able to predict for unseen data. Our dataset may have thousands of parameters or dimensions. In this case, the parameters will try to adjust with the noise in the data. Then, the training accuracy will be high and out of sample data gives low accuracy. Adding dropout to the data will assign zeros to a percentage of data. This will happen for each epoch. Adding drop out layers can reduce the overfitting in LSTM networks. The loss functions are used by the optimization algorithm in every epoch to update the weights in every epoch. To predict categories, we have specific loss functions in keras library. The Hyper parameter tuning includes tuning of batch size, epochs, learning rate, activation functions, dropout layers, number of neurons etc.

## RESULTS

This is an on-going work. The number of total records is 555957. A sample of 500 instances is taken from each class and a test set is generated. The validation score was 62.2 when tested with validation data. The accuracy of the model can be increased. The hyper-parameter tuning is going on to increase model's prediction accuracy. we have to evaluate the model with different parameters using Grid search process.

## CONCLUSION

In this work, the main focus is automatic classification of complaints and queries in internet forums or sites. The usual machine learning classification problems needs pre-defined features or manual intervention is needed to create features from dataset. The future plan is to optimize the model and maximize the accuracy. Many predictive modeling problems of sequence classification can be solved using this method.

## REFERENCES

1. Oguzhan Gencoglu, "Automatic Classification of Forum Posts: A Finnish Online Health Discussion Forum Case", EMBEC 2017, NBC 2017: EMBEC & NBC 2017pp 169-172J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.
2. Sumit bhatia, Cliff Brunk and Prasenjit Mitra, "Analysis and automatic classification of web search queries for diversification requirements", Proceedings of the American Society for Information Science and Technology Volume 49, Issue 1, Version of Record online: 24 JAN 2013.

3. Lorenzo A. Rossi, and Omprakash Gnawali, "Language Independent Analysis and Classification of Discussion Threads in Coursera MOOC Forums", 15th IEEE International Conference on Information Reuse and Integration (IRI 2014), At San Francisco, CA.
4. Bernard J. Jansen, and Danielle Booth, "Classifying Web Queries by Topic and User Intent", April 14–15, 2010, Atlanta, GA, USA
5. D. Irazú Hernández, Parth Gupta, Paolo Rosso, and Martha Rocha "A Simple Model for Classifying Web Queries by User Intent", January 2012.
6. Kristof Coussemant, and Dirk Van den Poel "Improving customer complaint management by automatic email classification using linguistic style features as predictors", Decision Support Systems 44 (2008) 870– 882.