# KNOWLEDGE MANAGEMENT SYSTEMS FOR PROBLEM DETECTION AND DIAGNOSIS

*A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of*

**Doctor of Philosophy**

*in*

**Computer Science and Engineering**

*by*


**SUMANA DE**

**Roll. No. 16IT1305, PHD**

**Registration No. - NITD/PhD/CS/2018/01038**

*Under the Supervision of*

**Dr. Baisakhi Chakraborty,**

**Associate Professor**

Department of Computer Science and Engineering,
National Institute of Technology, Durgapur, India






**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, NATIONAL INSTITUTE OF TECHNOLOGY DURGAPUR, WEST BENGAL-713209, INDIA**


**MARCH, 2022**

# Abstract

The research problem is to develop Knowledge Management Systems (KMS) for Problem detection and diagnosis. After detection of the problem, the KMS helps to provide solutions to those detected and diagnosed problems. Tacit knowledge resident in the brain is extracted through socialization and stored as explicit knowledge in a Knowledge-Base (KB) which may be used and re-used to solve a problem. Knowledge Management (KM) refers to techniques and practices required to identify, capture and reuse knowledge effectively to solve new problems. Access to the right knowledge at the right time via an efficient Knowledge Management System (KMS) helps in problem-solving, fault detection, disease diagnosis, organizational performance enhancement, faster knowledge search, and several other applications The main goal of KMS is to define and explain knowledge so that it can be communicated formally, allowing it to be reused.

Among the several different types of KMS, this thesis proposes a total of seven different kinds of KMS which acts as Knowledge Application Systems for solving several different problems in the field of disease detection, diagnosis and prediction, automobile fault detection, and environmental problem detection as assistant tools to help detect problems and solve them: i) Chronic Kidney Disease Prediction System (CKDPS), ii) Anemia Severity Detection System (ASDS), iii) Disease Detection System (DDS), iv) Excessive-Smoke-Emitting-Car Fault Diagnostic System (ESECFDS), v) Automobile Smoke Problem Solution System (ASPSS), vi) Car Fault Diagnosis System (CFDS), and vii) Air Quality Condition Detection System (AQCDS). Each of the KM systems has been developed with different algorithms and methodologies for problem detection and diagnosis in different domains and application areas. Three different methodologies or techniques have been applied to develop the KM systems - Case-Based Reasoning (CBR) Methodology, Machine Learning Techniques, and Natural Language Processing. In the field of health problems, CKDPS has been developed Using Machine Learning Technique to diagnose the problems of Chronic Kidney disease; ASDS has been developed using both Machine Learning Technique and Case-Based Reasoning (CBR) to diagnose Anemia; DDS has been developed using Machine Learning Technique to diagnose a set of diseases related to the liver and hepatic disorders, Heart disease, Diabetes, and Chronic Kidney disease. In the domain of automobile problems, ESECFDS has been developed using Case-Based Reasoning (CBR) Methodology to solve the problem of excessive

smoke emission from cars; ASPSS uses Machine Learning Techniques and Natural Language Processing to solve the problem of smoke emissions from automobiles; CFDS is developed using Case-Based-Reasoning to solve the problems related to car faults. In the field of environmental problems, AQCDS is developed using Machine Learning Technique to detect the exact air quality condition of an area in India. The systems are user-friendly query-response systems; users provide problem descriptions as queries to the systems and the system provides the solutions as the feedback results to the users accurately in a minimum time.

For the development of the proposed systems, each of the different datasets for each of the different systems are collected from different internet websites. The datasets used for experimentation have some numerable records. So, according to those datasets, to identify diseases, DDS and CKDPS provide 100 percent accuracy, precision, and recall; ASDS has a 92 percent accuracy rate, an 84 percent precision rate, and a 90 percent recall rate. To diagnose car faults the time complexity of the ESECFDS, in the Best case and Average Case, is Constant: O(1), in the Worst case, is Linear: O(N), where N is the input size; The accuracy, precision, and recall of ASPSS are 92.6 percent, 94 percent, and 93 percent, respectively in the prediction of dominant color and ASPSS performs the detection efficiently with O(N) time complexity in both best, average, and worst cases; CFDS will give 100 percent accuracy, precision, and recall rate if the CB of the system is updated till the current year; lastly, AQCDS provides 93 percent accuracy, 90 percent precision, 90 percent recall, and 89 percent F1 score to detect air quality condition. Using python module called 'big-O', the time complexity of the algorithms of the proposed systems are measured as Linear time.

In the future, when the proposed systems will be upgraded with bigger datasets, the accuracy, precision, and recall results may vary. The systems may not provide 100 percent accuracy, precision, and recall results in the future.