



# STUDENT RANK BASED PREDICTION SYSTEM USING DATA MINNING TECHNIQUES

Sai Bhavani . V , Om Prakash Samantray Mtech , (ph.D)(Assistant Professor)

Department of Computer Science and Engineering , Raghu Institute Of Technology , Visakhapatnam , AP , India .

**Abstract :** Data Mining and Artificial intelligence are the state of the art research in today's technology. The data mining techniques are employed in various fields for classification, association and prediction. We attempt to apply data mining techniques like CART (Classification and Regression Trees) algorithm, SVMs (support vector machine) algorithm, KNN (K-nearest neighbor) algorithm to predict the college in which a student can secure admission. College Predictor helps students to get an idea about the chances of admission in the college of their choice. In this project we confined our system to AP EAMCET/NEET Exams only. We can predict college based on the entrance exam hall ticket number and rank using this web portal. We implement these algorithms to classify and predict the colleges based on student rank and category. This system generates the top three list of the colleges on the basis of candidates rank and previous year seat allotments.

## 1. INTRODUCTION

### Data Mining

“A secret of success is to know something that nobody Else knows “ said by Aristotle Onassis Data mining is the process of extracting useful information from given data set i.e Data mining is an analytic process designed to explore data in search of consistent patterns and or systematic relationships between variables , and then to validate the findings by applying the detected patterns to new subsets of data . The ultimate

goal of data mining is prediction and predictive data mining is the most common type of data mining and one that has the most direct business applications.

**Classification :** It is a data analysis task , i.e , the process of finding a model that distinguishes data classes and concepts .Classification is the problem of identifying to which of a set of categories (sub populations ) , a new a observation belongs to , on the basis of a training set of data containing

observations and whose categories membership .

**Prediction / Regression:** Regression techniques can be adapted for prediction .  
Regression

Analysis can be to model the relationship between one or more independent variables . In data mining independent variables are attributes already known and repose variables are what we want to predict .Unfortunately , many real-world problems are not simply prediction . For example , the CART (Classification and Regression Trees ) decision tree algorithms , SVM algorithm and KNN algorithm can be used to build both classification trees ( to classify categorical response variables ) and regression trees ( to forecast continuous response variables ) .Neural networks too can create both classification and regression models .Types of regression methods .

- Linear Regression
- Multivariate Linear Regression
- Nonlinear regression
- Multivariate non linear Regression

### Association Rules :

It is rule in the data mining method for discovering interesting relations between variables in large databases. If is intended to identify strong rules discovered in data bases using some measures of interestingness. This rules-based approach also generates new has it analyses more data. The ultimate goal, assuming a large enough, dataset, is to help a machine mimic the human brain's feature extraction and abstract association capabilities from new uncategorized data. For example, the rule found in the sales data of a supermarket would indicate that if a customer buys onions and potatoes together, they are likely to also buy hamburger meat. Such information can be used as the decisions about marketing activities such as, e.g., promotional pricing or product placement.

In addition to the above example from market basket analysis association rules are employed today in many application areas including Web usages mining , intrusion detection , continuous production , and bioinformatics .

Today all the work at the time of admission of the students is done manually by link and paper which is very slow and consuming much efforts and time . In the modern world of technology , computer are effecting our lives in more ways than we probably are aware of computerized management maintaining information of an educational institute , colleges , other the list is endless . The main principle behind the need of college admissions in professional colleges like engineering would face lots of problems . Admission in engineering colleges in the state of Andhra Pradesh or any state is based upon common entrance test (CET ) and since more than 1.5 lakh seats are to be allotted in more than 200 engineering and medical colleges and over 35 different branches of engineering , for student belonging to many categories like open , home university , outside home university , reserved category (SC ,ST , OBC etc ) the problem becomes more serious and students struggles to understand which colleges they are likely to get admitted in , even after going through cut-off data of previous years . Most of the students fill it inconvenient because of not having awareness or guidance in filling incorrect options and fail to get admission . To minimize the stress of students get the list of all colleges in which they could get the admissions the click of a button , making the admission process fast and easy . Based on past years trends and data analysis , it predicts the EAMCET Rank-wise colleges for candidates . Candidate simply require entering the details asked by the AP EAMCET / NEET college predictor 2018 and it will give the list of the college based on candidate rank or score . However , the candidate must be aware that there are chances that results predicted by the tool may not be 100% accurately .

### RELATED WORKS

Osmanbegovic, Edin; Suljic, Mirza Et .al [1]  
Data mining has become a significant problem in recent years in a variety of sectors throughout the world, including the subject of education. In the academic realm, the necessity to automate various activities such as grading students,

enhancing student retention, testing students, forecasting student success, and administrative chores such as material optimization has prompted the use of data mining techniques and methodologies. Dineshkumar B Vaghela Et .al [2] Any organization's most valuable asset is data, which is then processed to generate usable information. Data mining techniques are frequently utilised to increase revenues and grow a company's operations. Many studies have been conducted in recent years using data mining techniques on educational data to enhance the educational system. Data mining can be beneficial for forecasting student achievement and finding groups of students that behave similarly. Sharayu N. BondeEt .al .[3] Educational data mining (EDM) is a relatively new trend in the data mining and knowledge discovery in databases (KDD) fields, focusing on extracting useful patterns and learning from educational data frameworks such as admissions frameworks, registration frameworks, course administration systems, and other frameworks that manage students at various levels of education. Ravi Kumar Rathore Et . al. [4] The proposed student prediction system is an effective method for categorising student data based on merit. When the quantity of students and the subject of concern for placements are both high, classifying students based on merit is a time-consuming job. We propose student data categorization utilising a variety of algorithms, including decision trees, SVMs, KNN, CART, and neural networks. Boumedyen Shannaq Et . al .[5] The goal of this study paper is to better analyse current trends in higher education systems from the outside in order to determine which elements may result in loyal pupils. The

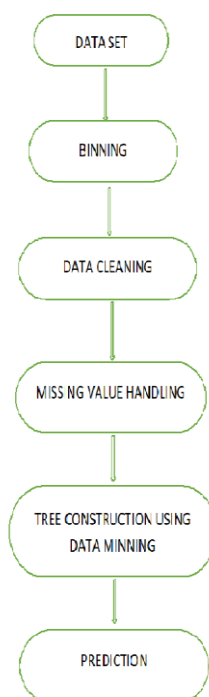
need for loyal students pushes higher education institutions to get to know them well; one method to achieve this is to use proper database administration and processing. Data mining technologies are a viable tool for extracting valuable information from current students in order to manage relationships with prospective students. This may suggest at an early stage which types of students will be enrolled and what areas in higher education systems should be focused on for assistance. Sellappan Palaniappan Et .al .[6] The dataset was made up of 18 features and one meta attribute, with 12 features coming from SIS, 2 from Moodle, four from eDify, and one result serving as the target feature for prediction. A total of 772 samples from a single academic year were included in the dataset. Eight classification methods generated from the literature study and other publications were used to evaluate the entire dataset. With an accuracy of 88.3 percent, the tree-based categorization model—specifically, consequence, the results obtained with fewer characteristics were superior than those obtained with all of the features. The CN2 Rule Inducer algorithm was the second-best performer, with an accuracy of 87.4 percent. The rationale for adopting CN2 Rule Viewer is that it offers rules induction with probability, which makes it easier to understand for non-expert users like faculty. TamayozEt .al . [7] In education, the prediction of student academic achievement has gotten a lot of attention. Despite the fact that learning outcomes are thought to promote learning and teaching, prognosticating student outcomes is still a work in progress. Academic achievement is carefully assessed using student learning outcomes, and a decade of study effort done between 2010 and

November 2020 was reviewed to offer a fundamental knowledge of the intelligent approaches utilised for the prediction of student performance. ACM, IEEE Xplore, Google Scholar, Science Direct, Scopus, Springer, and Web of Science were among the electronic bibliographic databases searched. Finally, we collected and evaluated 62 relevant publications, focusing on three aspects: (1) how learning outcomes are anticipated, (2) predictive analytics models designed to forecast student learning, and (3) the major elements influencing student results. To summarise and present the major findings, standard techniques for performing systematic literature reviews, such as PICO and PRISMA, were used. Learning outcomes were primarily assessed using performance class standings (i.e., rankings) and accomplishment scores (i.e., grades). Kalpesh P. Chaudhari Et . al [8] The level of education provided to students can be used to assess an academic institution's performance. Exploration of data pertaining to redirection regarding student performance is used to attain the best degree of quality in the educational system. The lack of a mechanism to analyse and assess pupils' performance and growth is not being addressed these days. There are two reasons why this occurs frequently. To begin with, the current system is ineffective in predicting student success. Second, pupils' performance is being harmed due to a lack of consideration of several critical factors. Because of the enormous amount of data in academic databases, predicting students' success is a more difficult process. This suggested approach can assist in more precisely predicting student performance. A appropriate data mining method will be used for them.

Ahmed MueenKingAbdulaziz Et . al .[9] The major goal of this research is to use data mining techniques to predict and assess students' academic performance based on their academic records and involvement in discussion forums. EDM (Educational Data Mining) is a new academic intervention technique. EDM can be used by educational institutions to conduct in-depth analyses of student characteristics. We gathered data from students in two undergraduate classes for this investigation. The dataset was classified using three distinct data mining techniques (Nave Bayes, Neural Network, and Decision Tree). Three classifiers' prediction performance is tested and compared. The Nave Bayes classifier outperforms the other two classifiers, with an overall prediction accuracy of 86 percent. This research will assist teachers in helping students enhance their academic performance. Elaf Abu Amrieh, ThairHamtni Et . al [10] In recent years, educational data mining has gotten a lot of attention. To extract hidden knowledge from educational data, a variety of data mining approaches have been developed. The information gathered aids institutions in improving their teaching techniques and learning processes. All of these enhancements contribute to improved student performance and educational outcomes. A collection of classifiers, including Artificial Neural Networks, Nave Bayesian, and Decision Tree, are used to evaluate the performance of the student's predictionmodel. In addition, to increase the performance of these classifiers.Huda Al-

Shehri Amani Al-Qarni Et . al .[11] This paper offered two prediction models for estimating a student's final test result. The study used a popular dataset supplied by the University of Minho in Portugal, which has 395 data samples and is related to arithmetic performance. Forecasting student performance might be beneficial for adopting early precautions, taking immediate action, or selecting a student who is suitable for a certain activity. The importance of developing better models in order to improve performance cannot be overstated. The K-Nearest Neighbor method was employed in most previous work on the same dataset, which yielded poor results, whereas the Support Vector Machine approach was rarely utilised, as is the case now.

### 3 .System Outline Architecture :



**Fig : 1 System Architecture :**

## Description

### Data Set :

This is a sample dataset we have used in this project as a training dataset. In this we have eleven types of attributes to identify a suitable college for a student according to their rank and details so, here we collected from previous year seat allotments to the perspective colleges we can predict the current students to get the chances of admissions by their rank and some required details. In this dataset Rank play a major role to predict the system with out having rank the student was not eligible to get seat in any college to any ware but, some students can join by management quota. This project is highly useful for the student who attempted the common entrance test for engineering students .

### Data Cleaning:

Real-world data tend to be incomplete, noisy, and inconsistent. *Data* cleaning(or data cleansing) routines attempt to fill in missing values, smooth out noise while identifying outliers, and correct inconsistencies in the data. In this section, you will study basic methods for data cleaning.

### Binning:

Binning or discretization is the process of transforming numerical variables into categorical counterparts. An example is to bin values for Age into categories such as 20-39, 40-59, and 60-79. Numerical variables are usually discretized in the modelling methods based on frequency tables (e.g., decision trees). Moreover, binning may improve accuracy of the predictive models by reducing the noise or non-linearity. Finally, binning allows easy identification of outliers,

invalid and missing values of numerical variables.

### **Treatment for Missing Values:**

Many existing, industrial and research data sets contain Missing Values. They are introduced due to various reasons, such as manual data entry procedures, equipment errors and incorrect measurements. Hence, it is usual to find missing data in most of the information sources used. The detection of incomplete data is easy in most cases, looking for Null values in a data set. However, this is not always true, since Missing Values (MVs) can appear with the form of outliers or even wrong data (i.e., out of boundaries)

### **Tree Construction:**

The tree will be constructed based on Cart Algorithm. Where the target variable is categorical and the tree is used to identify the "Class" within which a target variable would be fall into. Rules for splitting data at a node based on the value of one variable; Stopping rules for deciding when a branch is terminal and can be split no more and finally, a prediction for the target variable in each terminal node.

### **Prediction:**

Prediction in data mining is to identify data points purely on the description of another related data value. It is not necessarily related to future events but the used variables are unknown. Prediction derives the relationship between a thing you know and a thing you need to predict for future reference.

In this project we are predicting AP EAMCET and NEET students rank based colleges.

## **4.EXPERIMENT AND RESULTS**

**Sample data set :**In This project we collected previous year AP EAMCET and NEET counselling dataset and we gathered all the data into MS EXCEL spreadsheet and then we converted the entire data into .CSV file format. Under the dataset we mentioned ten columns with lots of records , the attributes are exam type, exam sub-type, gender, location, rank category, interested branch, inter stream, college admitted. In this project we used three different datasets, which are Top colleges in state of Andhra Pradesh, Training dataset and testing dataset under the testing dataset the college admitted column is not mentioned and reaming columns are same as training dataset. In this project we used hypothetical dataset.

This is a sample dataset we have used in this project as a training dataset. In this we have eleven types of attributes to identify a suitable college for a student according to there rank and details so, here we collected from previous year seat allotments to the perspective colleges we can predict the current students to get the chances of admissions by there rank and some required details. In this dataset Rank play a major role to predict the system with out having rank the student was not eligible to get seat in any college to any ware but, some students can join by management quota.

S.No	COLLEGE NAME	BRANCH NAME	BRANCH CODE	HALL TICKET NO	RANK	NAME OF THE CANDIDATE	SEX	CAST E	REGIO N
1	MAHAKAJ VIJAYARAM GAJAPATHI RAJ COLLEGE OF ENGINEERING	COMPUTER SCIENCE AND ENGINEERING	CSE	2021J01017	48015.00	MOHAMMED SAMEERUDDIN	M	BC_E	OU
2	MAHAKAJ VIJAYARAM GAJAPATHI RAJ COLLEGE OF ENGINEERING	COMPUTER SCIENCE AND ENGINEERING	CSE	2021O01057	36906.00	BANOTHU PRASHANTH	M	ST	OU
3	MAHAKAJ VIJAYARAM GAJAPATHI RAJ COLLEGE OF ENGINEERING	COMPUTER SCIENCE AND ENGINEERING	CSE	2021Q01465	55006.00	GURRAM MANNOJ	M	SC	SVU
4	MAHAKAJ VIJAYARAM GAJAPATHI RAJ COLLEGE OF ENGINEERING	COMPUTER SCIENCE AND ENGINEERING	CSE	2021O01057	36906.00	BANOTHU PRASHANTH	M	ST	OU
5	MAHAKAJ VIJAYARAM GAJAPATHI RAJ COLLEGE OF ENGINEERING	COMPUTER SCIENCE AND ENGINEERING	CSE	2021Q01465	55006.00	GURRAM MANNOJ	M	SC	SVU
6	MAHAKAJ VIJAYARAM GAJAPATHI RAJ COLLEGE OF ENGINEERING	COMPUTER SCIENCE AND ENGINEERING	CSE	2022J01402	68159.00	CHENNURI UMESH VARMA	M	SC	OU
7	MAHAKAJ VIJAYARAM GAJAPATHI RAJ COLLEGE OF ENGINEERING	COMPUTER SCIENCE AND ENGINEERING	CSE	2022O01218	47511.00	MARAGONI RUTHIK	M	BC_B	OU
8	MAHAKAJ VIJAYARAM GAJAPATHI RAJ COLLEGE OF ENGINEERING	COMPUTER SCIENCE AND ENGINEERING	CSE	2023A03759	22709.00	S UJWAL	M	BC_B	OU
9	MAHAKAJ VIJAYARAM GAJAPATHI RAJ COLLEGE OF ENGINEERING	COMPUTER SCIENCE AND ENGINEERING	CSE	2023B02100	77962.00	GORRE SAI RAJU	M	SC	OU
10	MAHAKAJ VIJAYARAM GAJAPATHI RAJ COLLEGE OF ENGINEERING	COMPUTER SCIENCE AND ENGINEERING	CSE	2023D02077	26893.00	ROHAN KULKARNI	M	OC	OU

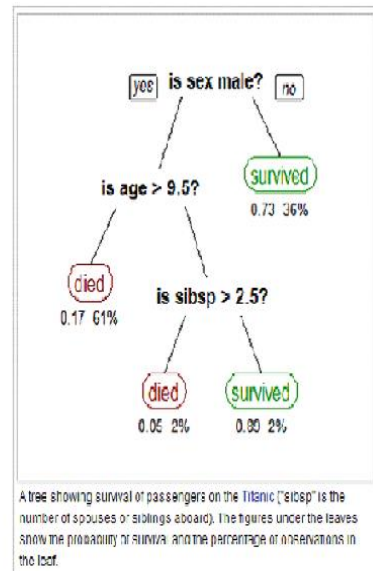


Fig 5 :Cart decision tree example

## SUPPORT VECTOR MACHINE (SVMs)

Support vector machine is a supervised algorithm. It can be used for both classification or regression challenges. Support Vector Machine algorithm is mainly used to solve classification problems. Support vectors are nothing but the coordinates of each data item. Support Vector Machine is a frontier that differentiates two classes using hyperplane. The goal of the svm algorithm is to create the best line or decision boundary that can segregate n – dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called hyperplane.

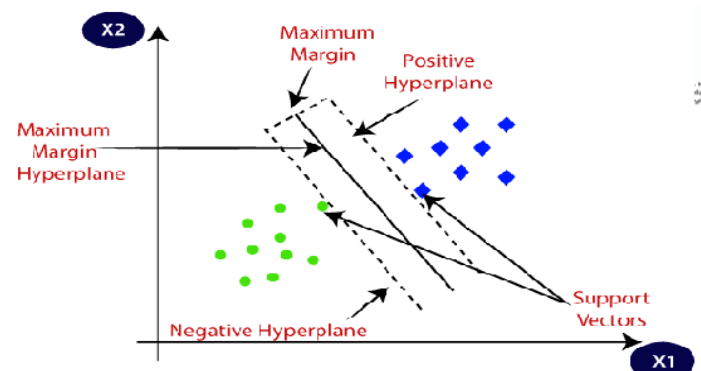


Fig : 6 SVM [Support Vector Machine]

### ALGORITHMS :

#### CART:

It stands for classification and regression trees and was introduced by Breiman in 1984. It builds both classifications and regression trees. The classification tree construction by CART is based on binary splitting of the attributes. It is also based on Hunt's algorithm and can be implemented serially. It uses Gini index splitting measure in selecting the splitting attribute.

Technology which point there are no more questions.

## K-NEAREST

## NEIGHBORALGORITHM(KNN)

KNN is a supervised learning algorithm. it is used for classification and regression predictive problems. In both cases the input consists of the closest training examples data set. The output depends on weather k-NN is used for classification and regression. The K-NN is a simplest of all data mining algorithms. we do not need

any explicit training. It can be used for both classification and regression.



**Fig 7** :KNN [ K- Nearest Neighbor Algorithm ]

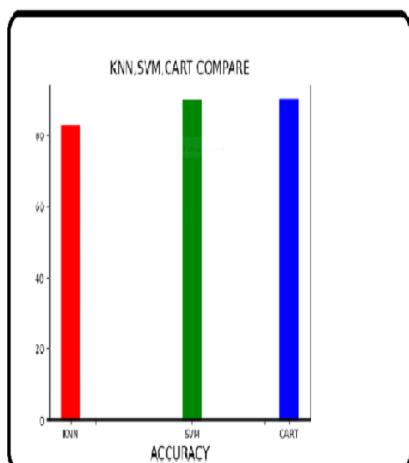
### Results :

ACCURACY	
KNN	80%
SVM	82%
CART	90%

The above table contains KNN has achieved an accuracy of 80 % ,SVM has achieved an accuracy of 82 % ,CART has achieved an accuracy of 90 % CART has outperformed then other 2 algorithms , Hence students selects the CART algorithm for better prediction .

This is the graphical representation for the candidate to go through with better understanding .

And, this system also grown thought with some of classification tasks for predicting the accuracy of system by a confusion matrix like Recall , precision and ROC accuracy for determining the accuracy of the system .



PREDICTION S	RECAL L	PRECISIO N	ROC
1	87.73%	94.12%	96.69 %
2	80.30%	90.09%	85.04 %
3	89.76%	87.85%	89.75 %



## 5 .Conclusion and Future

### Works

This project helps in making decisions regarding the colleges they should option based on their score, this had been developed for the candidates to know the various colleges in which they can get admitted based on their rank. This project predicts the colleges based on their rank and it will be able to make informed decision before the commencement of counselling and accordingly, it chooses the institute/University and programme in which they wish to take admission. This project works under Data Mining techniques. CART algorithm has the best accuracy as compare to SVM and KNN .

Data mining is an emerging and uncovered area for research work of Data mining. Data mining techniques can be useful in deriving patterns from educational data, and this pattern can be useful to improve Education System. This Project work can be helpful for Educational System. Security is the major concern, so in future the security parameters can also be considered.

### References :

- [1]Osmanbegovic, Edin; Suljic, Mirza (2012) : Data Mining Approach for Predicting Student Performance, Economic Review: Journal of Economics and Business, ISSN 1512-8962, University of Tuzla, Faculty of Economics, Tuzla, Vol. 10, Iss. 1, pp. 3-12Website :<https://www.econstor.eu> .com
- [2] Dineshkumar B Vaghela PhD Scholar Gujarat Technological University, Chandkh , Priyanka Sharma, Ph.D PhD Guide Gujarat

Technological University, Communications on Applied Electronics (CAE) – ISSN : 2394-4714 Foundation of Computer Science FCS, New York, USA Volume 2 – No.1, June 2015 – web site : [www.caeaccess.org](http://www.caeaccess.org)

[3] April 2018  
DOI:[10.1109/ICICCT.2018.8473228](https://doi.org/10.1109/ICICCT.2018.8473228)

Conference: 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)  
Project: [educational data mining](https://www.researchgate.net/publication/328111111) Sharayu N. Bonde website : [https :// www.researchgate.net.com](https://www.researchgate.net.com)

[4] Ravi Kumar Rathore<sup>1</sup>, J. Jayanthi<sup>2</sup>, Gurpreet Kaur Kaur lovely professional university Punjab, india International Journal for Research in Applied Science & Engineering Technology (IJRASET) website :<https://www.ieceeplore.com>

[5] 10 Issue 11 (Ver. 1.0) October 2010 Global Journal of Computer Science and Technology Boumedyen Shannaq ,Yusupov Rafael ,V. Alexandro WEBSITE <https://www.researchgate.net.org>.

[6] Raza Hasan<sup>1,2,\*</sup> , Sellappan Palaniappan<sup>1</sup> , Salman Mahmood<sup>1</sup> , Ali Abbas<sup>2</sup> , Kamal Uddin Sarker<sup>3</sup> and Mian Usman Sattar<sup>1</sup> Received: 4 May 2020; Accepted: 3 June 2020; Published: 4 June 2020 website :<https://www.ieceeplore.com>

[7] December 2020 Applied Sciences 11(237):1-28  
DOI:10.3390/app11010237 Project: Tamayoz 1 . website :[https:// www.ieceeplore.com](https://www.ieceeplore.com)

[8] Kalpesh P. Chaudhari, Riya A. Sharma<sup>2</sup>, Shreya S. Jha, Rajeshwari J. Bari

Amirah Mohamed Shahiria,\* ,  
WahidahHusaina , Nur'aini Abdul Rashida  
, aSchool of Computer Sciences  
UniversitiSainsMalayisa 11800 USM,  
Penang, Malaysia website :  
<https://sciencedirect.com>

[9] November 2016 International Journal of  
Modern Education and Computer  
Science 11(11):36-42

DOI:[10.5815/ijmeecs.2016.11.05](https://doi.org/10.5815/ijmeecs.2016.11.05) author :  
Ahmed MueenKing Abdulaziz University,  
Saudi Arabia,  
Jeddahwebsite:<http://www.meecs-press.org>

[10]Elaf Abu Amrieh, ThairHamtnini and  
Ibrahim  
September2016DOI:[10.14257/ijdta.2016.9.8.13](https://doi.org/10.14257/ijdta.2016.9.8.13)  
<https://www.computerresearch.com>

[11] Huda Al-Shehri Amani Al-QarniPublished  
in: 2017 IEEE 30th Canadian Conference on  
Electrical and Computer Engineering  
(CCECE)Date of Conference: 30 April-3 May  
2017 Date Added to IEEE Xplore: 15 June  
2017ISBN Information:INSPEC Accession  
Number: 16963480

