Formatted: Font: 5 pt, Complex Script Font: 9 pt

Formatted: (Complex) Arabic (Egypt), English (United States)

Formatted: Indent: Before: 2 cm

JAUES

Journal of Al Azhar University Engineering Sector

Vol. 13, No. 48, July 2018, ~~747~~894-~~769~~907

Formatted: Font: 8 pt, Complex Script Font: 8 pt

Formatted: Centered, Space After: 0 pt

A TEXT CLASSIFICATION APPROACH FOR EVALUATION DELAY CLAIMS

Formatted: Font: 14 pt, Complex Script Font: 14 pt

~~By~~ **Akram M. Hammam**[1]~~,~~[1], **Omar H. El-Anwar**[2]~~,~~[2]~~;~~ **and Moheeb El-Said**[3]

[1] Senior Construction Forensic Claims/ Delay Analyst and Ph.D. Candidate, ~~email:akramhammam@hotmail.com~~

[2] ~~Professor,~~ Structural Engineering Department, Construction Engineering and Mana Faculty of Engineering, Cairo University, ~~email: elsaid1204@yahoo.com~~

Formatted: Font: 12 pt, Complex Script Font: 12 pt

Formatted: Centered

Formatted: Font: 12 pt, Complex Script Font: 12 pt

Formatted: Font: 12 pt, Complex Script Font: 12 pt

Formatted: Space After: 0 pt

Formatted: Space After: 0 pt, Line spacing: Exactly 12 pt

ABSTRACT:

The significant rise in the complexity and scope of construction projects led introduction of highly advanced building systems characterizes the current con industry. This entails a significant increase in coordination and planning and ar change in the management culture of all project participants. On the other hand stakeholders are faced with an increasing demand from project owners to implem track programmes to achieve an early return on investment.

Consequently, claims and disputes throughout the majority of project delivery syste surged, influenced by the project parties' inability to effectively manage the claims The aim of this paper is to introduce a new methodology for the automatic text class of project delay claims documents to enhance efficiency in the management of del process. The proposed model utilizes activity and Work Breakdown Structure keyw given delay event activity path for the training of the proposed model, which is ther to predict unlabeled project documents. The proposed model has been implemen series of delay claims events in a mega project, the implementation yielded promisin in the performance evaluation measures (precision, recall, and F1-Score) compared t text classification models.

Formatted: Font: Bold

KEYWORDS: Delay Analysis, Text Mining, Claims, Data Mining, Naïve Bayes

Formatted: Font: Bold, Complex Script Font: Bold

INTRODUCTION

Formatted: Font: 12 pt, Complex Script Font: 12 pt

Disputes are arising between the Claimant and the Defendant after the failure to amicable settlement over unsettled claims. Recent studies revealed that the con disputes reported between 2012 and 2014 are taking a longer duration to resolve than

Formatted: Space Before: 0 pt, After: 0 pt, Line spacing: Exactly 12 pt

Formatted: Space After: 0 pt, Line spacing: Exactly 12 pt

[1] Senior Construction Forensic Claims/ Delay Analyst and Ph.D. Candidate, email:akramhammam@hotmail.com

[2] Project Controls Specialist, New York City Department of Design and Construction, and Affiliate Faculty of C Management, University of Washington, Seattle, WA, email: elanwar@uw.edu.

[3] Professor, Structural Engineering Department, Construction Engineering and Management, Faculty of Engine University, email: elsaid1204@yahoo.com

Formatted: Font: 8 pt, Not Bold, Complex Script Font: 8 pt, Bold

Formatted: Font: 1 pt, Complex Script Font: 1 pt

years and are ranging between 10-12 months (ARCADIS 2013; ARCADIS 201 prolonged duration is inter alia caused by the increased complexity of the current requiring review and assessment of the significant amount of related documents.
From the different categories of construction claims, delay and disruption claims ar the most complex to substantiate, requiring a well-established record-keeping process to substantiate the Claimant's rights to claim. In delay claims, the Claimant is rec demonstrate the nexus between the damages and the delay's event(s). Consequently,

Claimants may experience substantial losses due to failure to substantiate their claims the provision of sufficient evidence in a timely manner. (Braimah 2013; Carnell 2005) The challenges encountered during the delay claims process are attributed to the a substantiate and provide merit for the Claim. This crucial step constitutes one of exhaustive and time consuming tasks in substantiation of delay related claims, which the Claimant to correlate different types of project documents (such as Material Su (MS), Requests for Information (RFI), and Daily Site Records (DSR), etc.) for a delay event (Pickavance 2005).
In the last decade, and with the availability and low data storage costs, an average siz produces a vast amount of project documents of different types and formats. Hence, tl to extract relevant information becomes a challenge. However, with the introductio mining techniques, it became possible to extract relevant information (Soibelman 2002).
This paper introduces a new approach in the application of text mining in constructi claims. In the proposed approach, the information extracted from the project time sch used to train and classify project related documents related to a specified delay cla paper provides a brief description to the delay analysis methodologies used to apporti events then provides an overview of data/text mining and its importance in addressin vast amount of data. The paper then details the proposed methodology and its advanta the existing approaches in training the models using supervised learning, an introduces a real-life example of the proposed methodology.

Formatted: Space Before: 0 pt, After: 0 pt, Line spacing: Exactly 12 pt

Formatted: Space After: 0 pt, Line spacing: Exactly 12 pt

Background

Delay claims constitute a large portion of the construction claims and are considered common and complex form of construction disputes (Carnell 2005). Establishi claims mainly consists of three major components: (1) Establishing the factual inf and evidence that substantiate the Claimant's entitlement of additional time (st assessment); (2) Contractual evidence that support the Claimant's entitlement for a time; and (3) apportionment of delay claim (delay analysis) (Fawzy and El-adawa The four primary delay analysis methodologies are (1) As-Planned vs. As-Built; (2) I As Planned; (3) Collapsed As Built; and (4) Time Impact Analysis (Caletka 2008 ; and Linnett 2006). There are two major industry guidelines that are widely use reference for delay analysis methodologies; namely, the Society of Construction La and Disruption Protocol (SCL Protocol) and the Association for the Advancement Engineering International (AACEI) in the form of its 'Recommended Practice No. Forensic Schedule Analysis (RP-FSA)' (AACEI Commitee 2011).
In construction projects, the cost impact associated with establishing entitlement to d disruption claims is unanticipated during the tender phase and often results in a burden on the Claimant (Caletka 2008). One of the major challenges of esta entitlement in delay claims lies in the Claimant's efficiency in extracting relevan records in a timely and accurate manner. It can be observed from the ruling of n Court cases that the Claimants' failure to substantiate their claims was mainly attri their inability to provide contemporaneous records and evidence referring to cc (Wharf Properties Ltd v Eric Cumine Associates, 1991, The Foundation Co of Cana United Grain Growers Ltd, 1995 and Fru-Con Construction Corporation v The Unit 1999).

Formatted: Font: 8 pt, Not Bold, Complex Script Font: 8 pt, Bold

Formatted: Font: 1 pt, Complex Script Font: 1 pt

Even with well-documented projects supported by modern Document Management (DMS); the challenge remained in adapting those documents to produce a credible co within a fixed period of time and limited budget. It is established that the retrospectively proving a Claimant's entitlement could be very expensive dependin credibility of the gathered data (Pickavance 2005). Vital evidence required to subst claim may take prolonged periods of time to recognize and retrieve (Vidogah and I 1998). Consequently, claim experts estimate that 90% of arbitrator's time is cons establishing facts of a claim and if those facts are not demonstrated *unambiguously* t is anticipated to fail (Pickavance 2005).

Consequently, researchers and industry professionals recognized the need to esta efficient and accurate methodology to address those impediments resulting in si financial losses to organizations. This methodology should be able to accurately project records that are relevant to a specified delay event pertaining to a series of a Text Mining has the potential to achieve this objective as outlined in the subsequent which start by introducing text mining and distinguishing it from data mining; then c relevant performance measures, previous research using text mining in construction aim of this study and the research gap it is filling.

Formatted: Indent: Before: 0 cm, Hanging: 3.02 cm, Space After: 0 pt, Line spacing: Exactly 12 pt

Formatted: Space After: 0 pt, Line spacing: Exactly 12 pt

Text Mining

The last decade witnessed technological innovations in data storage technology substantial increase in storage capacities at an affordable cost; which in turn re accumulating more data across all industries (Bramer 2007). On the other hand, bu were not able to adapt to this surge; thus limiting the ability to examine and ex concealed knowledge (Chimay 2005). Accordingly, data and text mining methodolog developed to address this gap, as briefly introduced in the coming subsections.

Unlike the data mining process; data in text mining is unstructured. Hence, additio are required to formulate unstructured text content in a structured format to apply Learning (ML) algorithms. Text classification (TC), which forms an integral part Mining, applies a standard methodology consisting of ***the following steps***, (see Figure

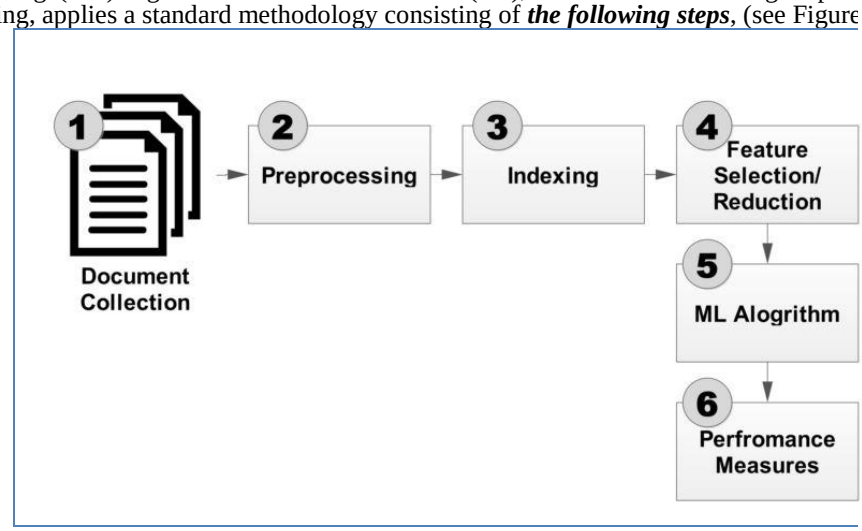

Figure 1111 – Text Mining Process

Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Not Bold

Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Not Bold

Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Not Bold

Formatted: Space After: 0 pt, Line spacing: Exactly 12 pt

The process starts with ***document collection*** involving the collection of relevant docu a document set. Step 2 ***Preprocessing*** step characterizes text mining from data minir responsible for adjusting the data as per the different database normalization forms c document vector for each document in the dataset. Preprocessing includes seve processes, such as tokenization, stop-word removal, and frequency calculation (Al C Kandil 2013). The second step - ***Indexing*** - transforms the text document from a

Formatted: Font: 8 pt, Not Bold, Complex Script Font: 8 pt, Bold

Formatted: Font: 1 pt, Complex Script Font: 1 pt

document to a document vector. A bag of words representation is the most wid document vector for its ease of use in document classification (Aggarwal and Zhai 2 and Soibelman 2007). The following step is ***Feature Selection/Reduction***, which filtering unnecessary words (features) that are irrelevant and do not support classification process. At Step 5, ***ML Algorithms*** can be applied to perform the Machine Learning Algorithms (Classification, Clustering, etc…). Final Step applying performance measures which evaluates the effectiveness of the applied mod Machine learning algorithms used in text classification are categorized as Superv Unsupervised Machine Algorithms. This paper focuses on Supervised Machine algorithms. Naïve Bayes is a common machine learning technique for text class which uses the probability theory for classification (Bayes Classifiers) and has prov reasonably accurate. However, it does not consider the number of occurrences of each a text document, which is essential in classifying any given document (Witten ar 2005).

Multinomial Naïve Bayes, introduced by McCallum and Nigam (1998), represents the of occurrences of terms in a document by a *bag of words (BoW)*. The documents in e are then trained as samples drawn from a multinomial word distribution. As a re conditional probability of each document given a class is simply a product of the pr of each observed word in the corresponding class. Other numerous common addressing text classification include Rocchio Algorithm, K-Nearest Neighbor, and Vector Machines (SVM) (Al Qady and Kandil 2013).

Classification tasks are categorized as 1) Binary/multi-class classification (a single each instance – mutually exclusive) and 2) Multi-label Classification, where each inst be associated with one or more classes (Tawiah and Sheng 2013).

Performance Measures

Measuring the performance of the proposed text classification system involve referencing the results of the proposed classification system with the output resul classification output is generated through the expertise and knowledge of human exp process to authenticate the relevance of the document is referred to as 'Gold Star 'Ground Truth' (Salama and El-Gohary 2013). The most widely used measures to as classification performance are (1) ***Accuracy,*** *which* measures the number of ti classifier makes the correct prediction (i.e. the percentage of documents classified c ***(2) Precision and Recall,*** which are used to measure classification effectiveness; and ***Score (F1-Measure),*** which combines Precision and Recall.

Text Mining Researches in the Construction Industry

Several studies were conducted on the application of text mining in different aspec construction industry. One study compared the performance of human extraction of and relations from contract documents with an automated method developed using Language Processing (NLP). This method used the Concept Relation Identificati Shallow Parsing (CRISP) technique (Al Qady and Kandil 2010). The results conclu both techniques yielded relatively similar results. Another recent study was conducte use of automatic text classifiers for classifying documents according to their corres group (such as project divisions, CSI Format, etc.) or using similarities among sem related documents under various conditions examining their performance (Al Q Kandil 2013). An Ontology-based text classification was used in a compliance approach (ACC) to automate the environmentally related textual documents (Zhou Gohary 2015). Whereas, another research developed a hybrid approach utilizing a sir clustering algorithm (unsupervised learning method) to cluster project documents up a text classifier is used to classify other documents (Al Qady and Kandil 2014).

Aim of study

Claimants incur significant losses in their pursuit to substantiate delay claim challenges to assimilate vast amount of data from different sources (islands of info and extracting credible information to their claim. Those challenges are time consur expensive, prone to errors, and result in inaccurate conclusions. The overarching go study is to enhance the accuracy and efficiency of the delay claim process and enh

Claimant's ability to retrospectively establish credible information to substantiate claim utilizing one of the widely used supervised learning algorithms (Multinomi Bayes) to train and classify unlabeled documents of a large size project.
As several studies suggest, supervised learning algorithms, require a comprehensive Set able to classify unlabeled documents. The accuracy of this supervised learning n governed mainly by the accuracy of the Training Set and its ability to provide an representation of the model it represents (Salama and El-Gohary 2013). Howev Training Sets are dependent on the expertise of human experts to classify data set proved to be expensive, time consuming and prone to errors (Al Qady and Kandil 201 The study proposes a new methodology for training the classifiers, which automatically generating the Training Set in lieu of a Training Set developed by classification in the form human experts labelling a series of documents to the re classes. Human Classifiers are characterized by their subjective findings, low e especially in large data sets. The proposed Training Set utilizes the keywords of a pr set of delay events, their associated path of activities, and hierarchical Work Br Structure (WBS) in a given time programme. Those keywords present a compr description of the delay event and are further utilized by the model to predict future i The following sections provide a brief description to the proposed schema implementation in a real-world case study.

Proposed text mining schema to classify delay claims

~~The proposed schema utilizes a new approach in training a model in supervised~~ mechanisms. Contrary to the traditional method for training models based on the human classifiers, the proposed schema follows a more efficient approach for model utilizing the information available in time programmes. Activity and WBS descrip considered to offer clear and concise keywords describing the scope of an activity.
In addressing delay claims retrospectively through Time Impact Analysis (TIA), a del is detected by its impact on the longest path or on the contract completion date ~~2Figure 2Figure 2Figure 2Figure 2~~ shows an example of a typical delay event introdu project programme and the impact of the delay event on the critical path.

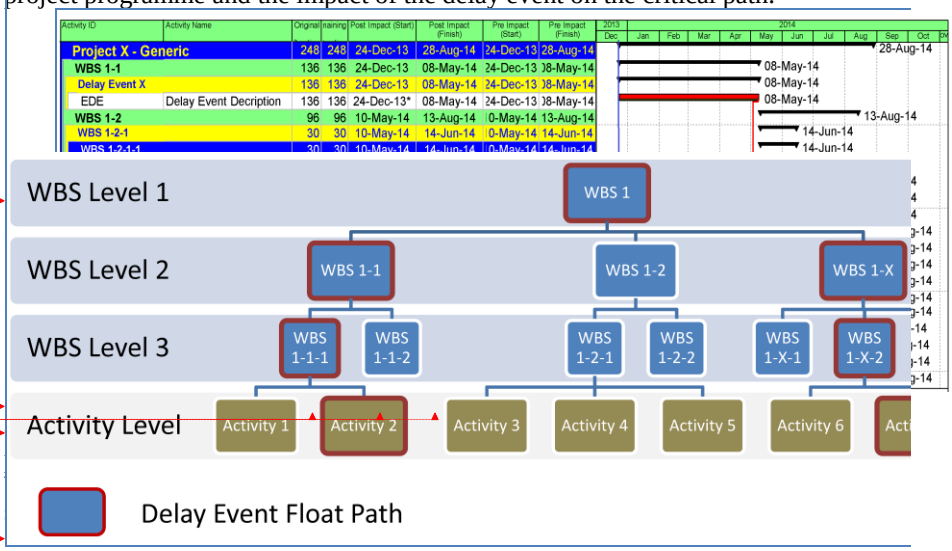

Figure ~~33333~~ – Selected WBS/Activity Elements forming Delay Event Path

The path of activities for all claim events is then tabulated in a structured format; thus the Training Set of the schema model (as demonstrated in Table 1).

Table ~~11111~~ – WBS Code/ WBS+Activity Keywords Forming the Training Set

Formatted: Font: 8 pt, Not Bold, Complex Script Font: 8 pt, Bold
Formatted: Font: 1 pt, Complex Script Font: 1 pt
Formatted: Font: 12 pt, Complex Script Font: 14 pt
Formatted: Space Before: 0 pt, After: 0 pt, Line spacing: Exactly 12 pt
Formatted: Space After: 0 pt, Line spacing: Exactly 12 pt
Formatted: Font: 12 pt, Not Bold, Complex Script Font: 12 pt, Bold
Formatted: Font: 12 pt, Not Bold, Complex Script Font: 12 pt, Bold
Formatted: Font: 12 pt, Not Bold, Complex Script Font: 12 pt, Bold
Formatted: Font: 12 pt, Not Bold, Complex Script Font: 12 pt, Bold
Formatted: Centered
Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Not Bold
Formatted: Space After: 0 pt, Line spacing: Exactly 12 pt
Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Not Bold
Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Not Bold
Formatted: Line spacing: Exactly 12 pt
Formatted: Space After: 0 pt, Line spacing: Exactly 12 pt
Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt
Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt
Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt
Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt
Formatted: Centered, Line spacing: Exactly 12 pt
Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt
Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt

WBS Code	WBS +Activity Keywords
WBS 1-2-1-1/Activity **1**	WBS 1-2-1-1 + Activity **1** Description Keywords
⋮	⋮
WBS 1-3-1-1-2-4/Activity **n**	WBS 1-3-1-1-2-4+ Activity **n** Keywords

A standard Multinomial Naïve Bayes algorithm coded using C# language is used to test the proposed model. The algorithm was tailored to integrate the time programme database table (Activity Description, WBS Keywords, etc.) in its structured format Multinomial Naïve Bayes algorithm to train the classifier. The proposed Training Se utilized to test a data set of documents to measure the effectiveness of the model. To the effectiveness of the proposed schema model, the model was applied to a real l project, which is demonstrated in the subsequent section

Methodology Implementation

In order to demonstrate the proposed schema capabilities and measure its effective proposed schema was applied to a mega project in the United Arab Emirates. A san delay claims was selected for the study with their associated project docume following subsection demonstrates the steps followed in the implementation of the p methodology, which is shown in Figure 4~~Figure 4Figure 4Figure 4Figure 4~~.

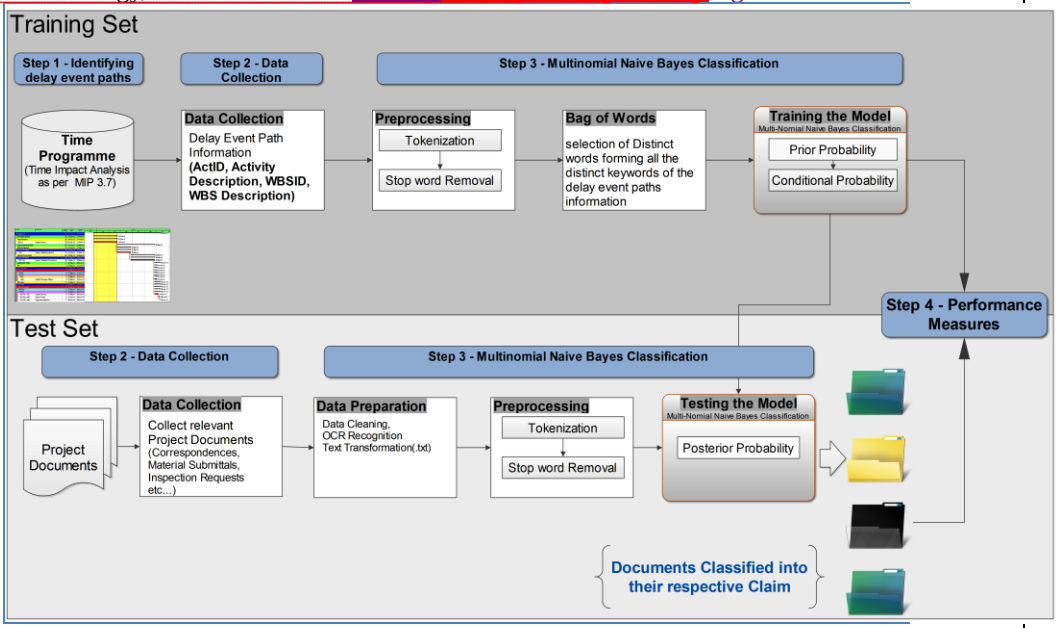

Figure ~~44444~~ – Proposed Schema Text Classification for Delay Claims

Step 1- Identifying the delay event paths

The Time Impact Analysis, which is recommended as per Method Implementation (MIP) 3.7 AACE industry guidelines (AACEI Commitee 2011), was used as t analysis methodology for the apportionment of the subject delay events for w following two steps were adopted:

Formatted: Font: 8 pt, Not Bold, Complex Script Font: 8 pt, Bold
Formatted: Font: 1 pt, Complex Script Font: 1 pt
Formatted: Line spacing: Exactly 12 pt
Formatted: Font: 12 pt, Complex Script Font: 14 pt
Formatted: Space Before: 0 pt, After: 0 pt, Line spacing: Exactly 12 pt
Formatted: Space After: 0 pt, Line spacing: Exactly 12 pt
Formatted: Font: 12 pt, Not Bold, Complex Script Font: 12 pt, Bold
Formatted: Font: 12 pt, Not Bold, Complex Script Font: 12 pt, Bold
Formatted: Font: 12 pt, Not Bold, Complex Script Font: 12 pt, Bold
Formatted: Font: 12 pt, Not Bold, Complex Script Font: 12 pt, Bold
Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Not Bold
Formatted: Space After: 0 pt, Line spacing: Exactly 12 pt
Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Not Bold
Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Not Bold

a. Each of the delay events was inserted in their respective updated window with data dates nearest to the start of the delay event. The delay events are then linke respective successor activities.

b. A path of activities comprising each delay event (DE) and the associate activities are identified and designated DE(X) for each delay event X in the Window.

Step 2- Data Collection

The document corpus (dataset) is comprised of (1) A Training Dataset – consist structured tabulated format of the keywords of the activity and WBS descriptions of of delay events of the subject claims, and (2) A Test Dataset consisting of the documents related to the delay events of the subject claims. The project documents letters, Minutes of Meeting (MOM), submittals, etc.

For the Training Set, the WBS and activities keywords forming the path of a activities of the delay path at a specified window of analysis are tabulated in a s format. The tabulated output forms a comprehensive set of descriptive keywo describe a delay event along the path. For example, in applying the method of Tim Analysis, the path of activities selected from the impact of delay event "*delay in the of Carpet colours*" consisted of a set of comprehensive keywords related to the del (Carpet - Material approval, Carpet Material – procurement, and Carpet flooring), as Table 2.

Table ~~22222~~ – WBS+Activity Keywords

ID	Activity	wbsCode/Activity	WBS+Activity Keywords	DE (Delay Event)
5668	PF-1370	ProjectTextMining.5.7.1.1 .1.4/PF-1370	ground level; security area;finishes;clean;snag	DE14
5669	PF-1440	ProjectTextMining.5.7.1.1 .1.4/PF-1440	ground level; security area;finishes;engineers inspection	DE14
5670	MT-1190	ProjectTextMining.8.10.4 /MT-1190	site establishment;mobilization;int ernal finishes;carpet	DE14
5671	PRC-ARC-140	ProjectTextMining.8.7.39 /PRC-ARC-140	prematerial procurement; internal finishes;carpet	DE14

For the Test Set, it is comprised of 54 project documents of different types (letters, Submittals (MS), Request for Information (RFI), etc.) pertaining to each of the eig claims listed in Table 3 which were previously classified by human experts. The pu the study is to test the efficiency of the proposed model when compared to the humai using a real-world example.

Table ~~33333~~ – Number of Documents in Each of the Eight Delay Claims

~~No.~~	~~Delay Event~~	~~No. of Docs~~
~~1~~	~~DE05—Delayed access to Security Area~~	~~8~~
~~2~~	~~DE08-Change in electrical scope of work~~	~~6~~
~~3~~	~~DE10-New substation power fed from power station~~	~~5~~
~~4~~	~~DE11-Handover of all new Toilet areas at Area 1~~	~~3~~

Formatted: Font: 8 pt, Not Bold, Complex Script Font: 8 pt, Bold
Formatted: Font: 1 pt, Complex Script Font: 1 pt
Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Bold
Formatted: Centered
Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Bold
Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Bold
Formatted: Space After: 0 pt, Line spacing: Exactly 12 pt
Formatted: Font: Times New Roman, 10 pt, Bold, Complex Script Font: 10 pt, Bold, English (United States)
Formatted: Normal, Centered

Formatted: Font: 8 pt, Not Bold, Complex Script Font: 8 pt, Bold

Formatted: Font: 1 pt, Complex Script Font: 1 pt

Formatted Table

No.	Delay Event	No. of Docs
~~5~~	~~DE14-Carpet Colours Selection~~	~~22~~
~~6~~	~~DE16-Revised signage design drawings - First Level~~	~~5~~
1	DE05 – Delayed access to Security Area	8
2	DE08-Change in electrical scope of work	6
~~8~~	~~DE61-Access to existing IAP area - Second Level~~	~~2~~
3	DE10-New substation power fed from power station	5
4	DE11-Handover of all new Toilet areas at Area 1	3
5	DE14-Carpet Colours Selection	22
6	DE16-Revised signage design drawings - First Level	5
7	DE17-Delayed access to LADA Area	3
8	DE61-Access to existing IAP area - Second Level	2

Formatted: Space After: 0 pt, Line spacing: Exactly 12 pt

Step 3 – Multinomial Naïve Bayes Classification

As previously mentioned, a standard Multinomial-Naïve Bayes Classification is ap
the proposed schema because of its efficient and ease of implementation (Rennie et a
A number of researchers considered the multinomial Naïve Bayes to be more effic
other Naïve Bayes Algorithms in document classification, especially in large diction
(McCallum and Nigam 1998). Unlike other Naïve Bayes algorithms, the document
the multinomial Naïve Bayes algorithm is determined by the number of word occ
(Chakrabarti et al. 2009; Stella and Faini 2009; Witten and Frank 2005).
The following section details the steps followed in applying the Multinomial-Naïv
(MNB) algorithm. The essential Data Preparation and Preprocessing Steps are impl
on the Training and Test Sets and then followed by the representation of each docu
applying the MNB algorithm (as shown in Figure 3).

Data preparation

The Training Set, as described earlier, is a structured table comprising of the
classifiers of the WbsCode and the Activity Id (wbsCode) and the Work Breakdown S
keywords (wbsKeywords) designated by the associated keywords of the corres
WbsCode and Activity Id. It is noted that the Training Set requires limited data pre
since the data is readily available in a structured format.
The selection of the keywords forming the delay event path creates a comprehensi
keywords characterizing the delay event. For example, in the path of driving activitie
by the delay event "Late Selection of Carpet Colours", the keywords associated
relevant activities and WBS keywords were found to be a good representation of t
event.
For the Test Dataset, Optical Character Recognition (OCR) software was utilized to
the printed text of project documents into encoded text (.txt) format to perform
mining classification process. It is worth mentioning that achieving a high accuracy
the transformation to text format is highly dependent on the quality of the original s
documents and scanning resolution (dpi). These conditions may not be fully achie
some real life projects; hence, achieving a high accuracy level is challenging a
compromise the accuracy of the output in such cases.

Pre-processing

Pre-processing of the documents corpus is an essential step in the text classification
To apply Data/Text mining techniques the data must be in a structured format. Hence
processing step transforms the documents to a structured format prior to the appli
data mining techniques and machine learning algorithms. The following are
processing steps implemented in the proposed schema model.
Tokenization – To apply data mining techniques on unstructured documents, it was
to breakdown the continuous characters into tokens such as words, sentences, etc. (
al. 2005). Tokenization was carried out on both the Training and Test Sets using an a

coded in C# Language, which involved breaking down the text into words sepa
spaces, as shown in Figure 5~~Figure 5~~Figure 5~~Figure 5~~Figure 5.

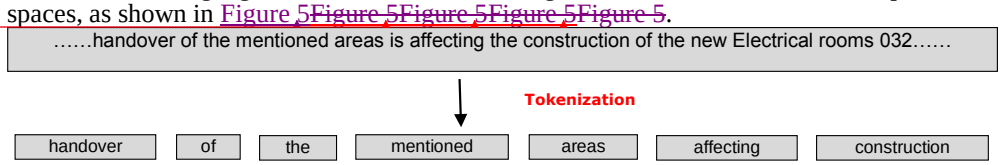

Figure ~~55555~~ – Tokenization Process

Stop Word Removal – This step involves the removal of words that are frequently c
and do not present significant importance to model training. A standard algorithm ir
utilized to remove words such as "the", "a", "and", etc., which fall into this category
as frequent words occurring in the specific document corpus as they do not imp
classifier's performance (Salama and El-Gohary 2013).
In the proposed schema, common stop words were removed from the bag of w
addition, stop words which were common in the document corpus such as letter h
titles providing a description of the document forms ("shop drawing submittal", "
Submittals", etc…) were considered stop words and were not significant to the cl
performance.
N-Grams – Part of the feature selection process is to select the relevant features (w
sequence of words) from the available data set to support model training. N-gra
process of characterization of n consecutive words in a document to provide a me
feature (Witten and Frank 2005). Bi-gram is a two-word feature used when two cor
words are used to better describe the feature, such as "board ceiling", "carpet flooring
In the Training Set, bi-grams and tri-grams were assigned as two- and three-word fe
the Bag of Words (BoW). Keywords such as '1st coat paint', 'board ceiling'
flooring', 'ceiling closure' are word representations forming all distinct features in th
Words. This approach induced further accuracy to the Training Set, as shown in Table

Table ~~44444~~ – Extract from the Vocabulary of Words

bagOfWordsWbsKeyword
1st coat paint
1st fix
2nd fix
3rd fix
board ceiling
carpet flooring
ceiling closure
ceiling grid
chiller
chiller plant
electrical room
hvac duct
mv cable
power on
sanitary ware
toilet accessories
toilets
ups room
wiring

Document Representation (BoW)

Formatted: Font: 8 pt, Not Bold, Complex Script Font: 8 pt, Bold
Formatted: Font: 1 pt, Complex Script Font: 1 pt
Formatted: Font: 12 pt, Not Bold, Complex Script Font: 12 pt, Not Bold, (Asian) Chinese (Simplified, PRC)
Formatted: Font: 12 pt, Not Bold, Complex Script Font: 12 pt, Not Bold, (Asian) Chinese (Simplified, PRC)
Formatted: Font: 12 pt, Not Bold, Complex Script Font: 12 pt, Not Bold, (Asian) Chinese (Simplified, PRC)
Formatted: Font: 12 pt, Not Bold, Complex Script Font: 12 pt, Not Bold, (Asian) Chinese (Simplified, PRC)
Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Bold
Formatted: List Paragraph, Space After: 0 pt, Line spacing: Exactly 12 pt
Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Bold
Formatted: Font: Not Italic, Complex Script Font: Not Italic
Formatted: Font: Not Italic, Complex Script Font: Not Italic
Formatted: Font: Font color: Auto
Formatted: Font: Not Italic, Complex Script Font: Not Italic
Formatted: Font: Not Italic, Complex Script Font: Not Italic
Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Bold
Formatted: Centered
Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Bold
Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Bold
Formatted Table
Formatted: Font: 12 pt, Complex Script Font: 12 pt
Formatted: List Paragraph, Justified, Indent: Before: 0 cm, Space After: 0 pt, Line spacing: Exactly 12 pt

Formatted: Font: 8 pt, Not Bold, Complex Script Font: 8 pt, Bold

Formatted: Font: 1 pt, Complex Script Font: 1 pt

Formatted: Space After: 0 pt, Line spacing: Exactly 12 pt

The plain text of the corpus documents is transformed to instances of fixed fea restricting the representation to the words only occurring in the Training Set. The dc are transformed into a Bag of Words (BoW). The BoW is a word documen representing the distinct words in the data set in which the rows are represe Documents (D) and the columns are represented by distinct features (F) of the word in the Training Set, as shown in Table 5 and Table 6.

Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Bold

Formatted: List Paragraph, Centered, Space After: 0 pt, Line spacing: Exactly 12 pt

Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Bold, Check spelling and grammar

Table ~~55555~~ – Extract of Term Frequency Matrix (Training Set)

wbsCode	1st fix	2nd fix	F032	north side	busbar
ProjectTextMining.5.2.SS-PF32-1.2.2.1.3/PF32-A-PH0-1350	0	0	0	1	0
ProjectTextMining.5.2.SS-PF32-1.2.2.1.3/PF32-A-PH0-1430	0	0	0	0	0
ProjectTextMining.5.2.SS-PF32-1.2.2.1.3/PF32-A-PH0-1490	0	0	0	1	1
ProjectTextMining.5.7.1.SS-PF32-1.1.4/PF33a-UD-POL-1370	0	0	0	0	0
ProjectTextMining.5.7.1.SS-PF32-1.1.8.1/PF33a-UD-POL-1210	0	0	1	0	0
ProjectTextMining.5.7.1.SS-PF32-1.1.8.1/PF33a-UD-POL-1220	1	1	1	0	0
ProjectTextMining.5.7.1.SS-PF32-1.1.8.1/PF33a-UD-POL-1280	0	0	1	0	0
ProjectTextMining.5.7.1.SS-PF32-1.1.8.1/PF33a-UD-POL-1290	0	0	1	0	0
ProjectTextMining.5.7.1.SS-PF32-1.1.8.2/PF33a-UD-POL-1410	0	0	1	0	0
ProjectTextMining.5.7.1.SS-PF32-1.1.8.2/PF33a-UD-POL-1470	0	0	1	0	0
ProjectTextMining.5.7.1.SS-PF32-11.1.2/PF33-UD-IMM-1120	1	0	0	0	0
ProjectTextMining.5.7.1.SS-PF32-11.1.2/PF33-UD-IMM-1140	1	0	0	0	0

Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Bold

Formatted: List Paragraph, Centered, Space After: 0 pt, Line spacing: Exactly 12 pt

Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Bold, Check spelling and grammar

Table ~~66666~~ – Extract of Term Frequency Matrix (Test Set)

DocID	DocName	2nd fix	F032	side	busbar	busbars	cabling	carpet	ceiling	chiller	chillers
1	DE05 140115 LC-FGF585-L-J0196	0	0	0	0	0	0	0	0	0	0
2	DE05 140206 TP01-SCR-0237 Rev.00	0	2	0	0	0	0	0	2	0	0
3	DE05 140222 LC-FGF585-L-J0263	0	5	0	0	0	0	0	0	0	0
4	DE05 140303 TP01-SCR-0276 Rev.00	0	2	0	0	0	0	0	0	0	0
5	DE05 140414 TP01-SCR-0358 Rev.00	0	0	0	0	0	0	0	3	0	0
6	DE05 140501 LC-FGF585-L-J0445	0	0	0	0	0	0	0	1	0	0
7	DE08 131024 TP01-MS-0008 Rev.01	0	0	0	0	0	0	0	0	0	0
8	DE08 131028 LPO Busbar	0	0	0	4	0	0	0	0	0	0
9	DE08 131116 TP01_YU65- MS- 0008-rev01-COM	0	0	0	0	0	0	0	0	0	0
10	DE08 131214-LC-FGF585-L-J0154	2	0	1	7	0	2	0	8	4	1
11	DE10 131215 TP01-SCR-0166 Rev.00	0	0	0	0	0	0	0	0	1	0
12	DE08 131222-T19-0200S-TP01-SL 0174	0	0	0	0	0	1	2	0	0	2
13	DE10 131224 TP01_YU65-CVI-CK-0007-rev00	0	0	0	0	0	0	0	0	0	0
14	DE10 140113 TP01_YU65-SCR-0166-rev00-COMM	0	0	0	0	0	0	0	0	1	0
15	DE10 140120 LC-FGF585-L-J0211	0	0	1	0	0	6	0	0	0	1
16	DE10 140127 TP01_YU65-ASR-MEP-0012-rev01-COMM	0	0	2	0	0	0	0	0	0	0

Formatted: Space After: 0 pt, Line spacing: Exactly 12 pt

Feature Selection

Limiting the features results in dimensionality reduction to the feature space, which i the efficiency of the model. For the training dataset, the keywords extracted f activity/WBS description are limited and comprehensive and did not require any reduction. As for the test dataset, stop word removal and N-grams were used fo reduction.

Text Classification algorithm

For training the model, a standard multinomial Naïve Bayes algorithm is applied base probability model formulated by Thomas Bayes (1701-1761). A tailored-made algori developed by the authors using C# Language to embed the Training Set and further the documents. The following steps were implemented:

- The Training Set, which includes the activities and Work Breakdown S forming the delay events, was exported from Primavera P6 software to a Microsof database and accessed through OLE DB (a COM-based application programming (API) for accessing data).
- For training the data corpus, the Prior Probability was computed, in w activities (wbsCode/Activity) forming the delay events' path were selected as the clas using Equation (1), where N_{c_j} is the Total number of activities (wbsCode/Activity) of N_c is the Total number of all activities (wbsCode/Activity).

$$p(c_j) = \frac{N_{c_j}}{N_c} \quad \text{(1)}$$

- Conditional Probabilities are then calculated using Equation (2) for each wo Vocabulary of words (as shown in Table 4). A parameter ($\propto$) for additive smoothing Smoothing $\propto = 1$) was used to avoid encountering zero probabilities $p(x_i|c_j)$ when u Equation (2), where $\mathbf{x_i}$ is the Word from the feature vector of a particular sample, $\mathbf{N}($ the Number of times feature $\mathbf{x_i}$ appears in samples from class c_j, $\mathbf{N}(\mathbf{c_j})$ is the term fre (number of counts) in the Training Set for class c_j, $\propto$: An additive smoothing parame for Laplace smoothing), **d**: Number of exclusive words (Vocabulary).

$$p(x_i|c_j) = \frac{N(x_i, c_j)+\propto}{N(c_j)+\propto d} \quad \text{(2)}$$

After determining the Prior and Conditional probabilities of the Training Set, the utilized to classify the documents of the Test Set using the posterior probability, as s Equation (3), where $p(c_j)$ is the Prior probability of a document occurring in class c_j, is the Conditional probability of feature x_i occurring in a document of class c_j, Number of documents in the Test Set.

$$p(c_j|d) = p(c_j)\prod_{k\ 1}^{m} p(x_i|c_j) \quad \text{(3)}$$

Step 4 - Performance Evaluation

To measure the effectiveness of the model, the following performance measures wer out.

Accuracy

Accuracy measures the number of documents in which the classifier makes the prediction (delay event) (i.e. the percentage of documents classified correctly) (as s Equation 4), where t_p is the True positives for class c_j, t_n is the True negatives for c is the Total number of documents in the Test Set.

$$Accuracy = \frac{t_p + t_n}{N} \quad \text{(4)}$$

Precision and Recall

Precision is the number of correctly retrieved document occurrences (t_p) by the model for the subject delay event divided by the total number of correctly documents and irrelevant retrieved documents to the subject delay event ($t_p +$ measure identifies reliability of the model to correctly predict relevant documents(as s Equation 5), where t_p is the true positives for class c_j, f_n is the False negatives for cl is the False positives for class c_j.

Whereas, **Recall** is defined as the number of correctly retrieved document occurrence the applied model for the subject delay event divided by the total number of retrieved documents and documents that are relevant documents to the subject delay e the model was not able to retrieve ($t_p + f_n$) (as shown in Equation 6). This measure i the model's ability to retrieve all the documents regardless of its relevance to the subj event.

$$Precision = \frac{t_p}{t_p + f_p} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (5)$$

$$Recall = \frac{t_p}{t_p + f_n} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (6)$$

F1-Score (F1-Measure)

F1-Score is a measure of the model effectiveness. The measure is regarded as the h mean between Precision and Recall (Jiansong Zhang 2016). Moreover, it is widely assess text classification systems and is calculated as shown in Equation (7) $Precision$ is the Measure calculated from equation (5), $Recall$ is the Measure ca from equation (6). F1-score values equal to 100% represents the perfect classifier.

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \ldots \ldots \ldots \ldots \ldots \ldots \ldots (7)$$

The proposed model was applied to the 54 project documents classified previously claims. To authenticate the 'Gold Standard' of the model performance measur computed and results are shown in Table 7. The proposed model achieved an accuracy of 99.1%, precision of 92.7% and recall of 93.3%.

The results outperform similar multi-class text classification approaches using au classification for training documents (Al Qady and Kandil 2013; Caldas et al. 2002; Z El-Gohary 2015). However, it should be noted that each study used a different ap example. It was also observed that the model's accuracy and precision measures decr delay events yielding similar spatial information, occurring at the same timing, and type of work like delay events DE08 (Change in Electrical Scope) and DE10 (New su power fed from power station). Similarly, delay events which pertain to an overall in all areas of the project cannot be classified, such as DE11 (Handover of all new Toile Area 1). This issue is referred to as multi-label classification, in which a document more than one class and it is not addressed in this paper.

Formatted: Font: 8 pt, Not Bold, Complex Script Font: 8 pt, Bold

Formatted: Font: 1 pt, Complex Script Font: 1 pt

Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Bold, English (United Kingdom)

Formatted: Centered, Space After: 0 pt, Line spacing: Exactly 12 pt

Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Bold, English (United Kingdom)

Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Bold, English (United Kingdom)

Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Bold, English (United Kingdom)

Formatted: Font: 10 pt, Bold, Complex Script Font: 10 pt, Bold, English (United Kingdom)

Formatted: Centered

Table 7~~7777~~ – Summarized Performance Measures of applied model

		Predicted																					
DE	Delay Event Description	01	02	03	04	05	06	07	08	Total Actual	t_p	t_n	f_p	f_n	Accuracy	Precision	Recall	F1-score
Actual Class																							
01	DE05 -Delayed access to Security Area	**8**	0	0	0	0	0	0	0	**8**	8	46	0	0	100.0%	100.0%	100.0%	100.0%
02	DE08-Change in electrical scope of work	0	**4**	0	0	0	0	1	0	**5**	4	49	0	1	98.1%	100.0%	80.0%	88.9%
03	DE10-New substation power fed from power station	0	0	**6**	0	0	0	0	0	**6**	6	48	0	0	100.0%	100.0%	100.0%	100.0%
04	DE11-Handover of all new Toilet areas at Area 1	0	0	0	**2**	0	0	0	1	**3**	2	51	0	1	98.1%	100.0%	66.7%	80.0%
05	DE14-Carpet Colours Selection	0	0	0	0	**23**	0	0	0	**23**	23	31	0	0	100.0%	100.0%	100.0%	100.0%
06	DE16-Revised signage design drawings - First Level	0	0	0	0	0	**4**	0	0	**4**	4	50	0	0	100.0%	100.0%	100.0%	100.0%
07	DE17-Delay Access to LADA Area	0	0	0	0	0	0	**3**	0	**3**	3	50	1	0	98.1%	75.0%	100.0%	85.7%
08	DE61-Access to existing IAP area - Second Level	0	0	0	0	0	0	0	**2**	**2**	2	51	1	0	98.1%	66.7%	100.0%	80.0%
	Total Predicted	**8**	**4**	**6**	**2**	**23**	**4**	**4**	**3**						99.1%	92.7%	93.3%	91.8%

Formatted: Space Before: 0 pt, After: 0 pt, Line spacing: Exactly 12 pt

Formatted: Font: 8 pt, Not Bold, Complex Script Font: 8 pt, Bold

Formatted: Font: 1 pt, Complex Script Font: 1 pt

Formatted: Font: 12 pt, Complex Script Font: 14 pt

Formatted: Space After: 0 pt, Line spacing: Exactly 12 pt

CONCLUSION

The cost and time associated with the settlement of construction claims have witn immense increase. A considerable percentage of Claimants encounter impedir substantiate their claims. Those impediments are not only limited to maintaining keeping system but also extracting the relevant documents pertaining to specific dela remain a major challenge. The current complexities of building systems and the si rise in the volume of data stored in current construction projects increase the challen task. These challenges necessitated the development of a more efficient class methodology.

This paper proposed a text mining classification methodology in delay claims u keywords of the activities and WBS of a group of delay event paths generat retrospective Time Impact Analysis (TIA) schedules. Those keywords forming pa Training Set are then used to predict and classify new project documents in the D Corpus. A Traditional Multinomial Naïve Bayes machine learning algoritl implemented for text classification. The performance measures of accuracy, rec precision yielded excellent results in classifying the project documents with their re delay event claims under study.

Formatted: Font: 8 pt, Not Bold, Complex Script Font: 8 pt, Bold

Formatted: Font: 1 pt, Complex Script Font: 1 pt

Formatted: Font: 12 pt, Complex Script Font: 12 pt

Formatted: Space Before: 0 pt, After: 0 pt, Line spacing: Exactly 12 pt

Formatted: Indent: Before: 0 cm, Hanging: 1.16 cm, Space Before: 0 pt, After: 0 pt, Line spacing: Exactly 12 pt, Numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Right + Aligned at: 1.9 cm + Tab after: 2.54 cm + Indent at: 2.54 cm, Tab stops: 0.97 cm, List tab + Not at 2.54 cm

Formatted: Font: Times New Roman, 12 pt, Complex Script Font: 12 pt, Do not check spelling or grammar

Formatted: Font: Times New Roman, 12 pt, Complex Script Font: 12 pt, Do not check spelling or grammar

Formatted: Font: Times New Roman, 12 pt, Complex Script Font: 12 pt, Do not check spelling or grammar

Formatted: Font: Not Italic, Complex Script Font: Not Italic

…matted

…matted

…matted

…matted

…matted

…matted

…matted

…matted

…matted

…matted

…matted

…matted

…matted

…matted

…matted

…matted

…matted

…matted

…matted

REFERENCES

1. AACEI Commitee. (2011). *AACE International Recommended Practice No. (FORENSIC SCHEDULE ANALYSIS)*, AACE International, West Virginia.
2. Aggarwal, C. C., and Zhai, C. (2012). *Mining Text Data*, Springer,New York.
3. Al Qady, M., and Kandil, A. (2010). "Concept Relation Extraction from Con Documents Using Natural Language Processing." J. Constr. Eng. M 136(3), 294 - 302.
4. Al Qady, M., and Kandil, A. (2013). "Automatic Classification of Project Docu the Basis of Text Content." J. Comput. Civ. Eng., 29(3), 1-11.
5. Al Qady, M., and Kandil, A. (2014). "Automatic clustering of constructio documents based on textual similarity." Autom. Constr., 42, 36–49.
6. ARCADIS. (2013). Global Construction Disputes 2013: A Longer Resolu Harris / ARCADIS, Amsterdam, Netherlands
7. ARCADIS. (2014). Global Construction Disputes 2014 - Getting the basi ARCADIS, Amsterdam, Netherlands
8. Braimah, N. (2013). "Construction Delay Analysis Techniques - A Re Application Issues and Improvement Needs." Buildings., 3(3) 506-531.
9. Bramer, M. (2007). Principles of Data Mining, Springer,New York.
10. Caldas, C. H., Soibelman, L., and Han, J. (2002). "Automated Classific Construction Project Documents." J. Comput. Civ. Eng., 16(4), 234-243.
11. Caletka, P. J. (2008). Delay Analysis in Construction Contracts, Wiley-B United Kingdom.
12. Carnell, N. J. (2005). Causation and Delay in Construction Disputes (Second Blackwell Publishing, United Kingdom.
13. Chakrabarti, S., Nadeau, T. P., Cox, E., and Neapolitan, R. E. (2009). Data Know It All, Elsevier, United Kingdom.
14. Chimay J. Anumba, C. O. (2005). Knowledge Management in Construction B Publishing, United Kingdom.
15. Fawzy, S. A., and El-adaway, I. H. (2013). "Contract Administration Guide Effectively and Efficiently Applying Different Delay Analysis Techniqu World Bank–Funded Projects." J. Leg. Aff. Dispute Resolut Eng. Constr., 4(2)
16. Jiansong Zhang, Nora M. El-Gohary. (2016). "Extending Building Informatio Semiautomatically Using Semantic Natural Language Processing Techni Comput. Civ. Eng., 30(5), 2246-2253.
17. Lin, K.-Y., and Soibelman, L. (2007). "Knowledge-Assisted Retrieval of Online Information in Architectural/Engineering/Construction." J. Constr. Eng. M 133(11), 871-879.
18. Lowsley, S., and Linnett, C. (2006). About Time- Delay Analysis in Cons RICS Books, United Kingdom.
19. Lucio Soibelman, M., and Kim, H. (2002). "Data preparation process for con knowledge generation through Knowledge Discovery in Databases." J. Comp Eng., 16(1), 39-48.
20. McCallum, A., and Nigam, K. (1998). "A comparison of event models for nai text classification." Dimension Contemporary German Arts and Letters, 752(1)
21. Pickavance, K. (2005). Delay and Disruption in Construction Contracts, 3rd I Professional Publishing, London - Singapore.
22. Rennie, J. D., Shih, L., Teevan, J., and Karger, D. R. (2003). "Tackling Assumptions of Naive Bayes Text Classifiers." Proc Int Conf on Machine L Artificial intelligence Laboratory; Massachusetts Institute of Technology, 20 623.
23. Salama, D. M., and El-Gohary, N. (2013). "Semantic Text Classification for Su Automated Compliance Checking in Construction." J. Comput. Civ 10.1061/(ASCE)CP.1943-5487.0000427,B4015001.

24. Stella, D. M., and Faini, M. (2009). "A Software System for Topic Extrac Document Classification." Proc. WI-IAT Workshops 2009., IEEE, Milan, Italy
25. Tawiah, C. A., and Sheng, V. S. (2013). "Empirical Comparison of Mu Classification Algorithms." Proceedings of the Twenty-Seventh AAAI Confe Artificial Intelligence., Association for the Advancement of Artificial Inte 10.1109/WI-IAT.2009.49
26. Vidogah, W., and Ndekugri, I. (1998b). "Review of the role of information tec in construction claims management." Comput. Ind., 35(1), 77–85.
27. Weiss, S. M., Indurkhya, N., Zhang, T., and Damerau, F. J. (2005). Text I Predictive Methods for Analyzing Unstructured Information., Springer, New Y
28. Witten, I. H., and Frank, E. (2005). Data Mining - Practical Machine Learni and Techniques, Second Edition., Elsevier, United Kingdom.
29. Zhou, P., and El-Gohary, N. (2015). "Ontology-Based Multilabel Text Classifi Construction Regulatory Documents." J. Comput. Civ. 10.1061/(ASCE)CP.1943-5487.0000346, 04015014.

Formatted: Font: 8 pt, Not Bold, Complex Script Font: 8 pt, Bold
Formatted: Font: 1 pt, Complex Script Font: 1 pt
Formatted: Font: Not Italic, Complex Script Font: Not Italic
Formatted: Font: Not Italic, Complex Script Font: Not Italic
Formatted: Font: Times New Roman, 12 pt, Complex Script Font: 12 pt, Do not check spelling or grammar
Field Code Changed
Formatted: Font: Times New Roman, 12 pt, Complex Script Font: 12 pt, Do not check spelling or grammar
Formatted: Font: Not Italic, Complex Script Font: Not Italic
Formatted: Font: Not Italic, Complex Script Font: Not Italic
Formatted: Font: 12 pt, Complex Script Font: 12 pt, Do not check spelling or grammar
Formatted: Font: 12 pt, Complex Script Font: 12 pt, Do not check spelling or grammar
Formatted: Font: Times New Roman, 12 pt, Complex Script Font: 12 pt, Do not check spelling or grammar
Formatted: Font: 12 pt, Complex Script Font: 12 pt