**LITERATURE SURVEY**

**1) Spam Review Detection Techniques: A Systematic Literature Review**

**AUTHORS:**  **Naveed Hussain 1,2, Hamid Turab Mirza 1,\*, Ghulam Rasool 1, Ibrar Hussain 2 and Mohammad Kaleem 3**

We consider the problem of planning the ISS cosmonaut training

with diﬀerent objectives. A pre-deﬁned set of minimum qualiﬁcation

levels should be distributed between the crew members with minimum

training time diﬀerences, training expenses or a maximum of the train-

ing level with a limitation of the budget.

First, a description of the cosmonaut training process is given.

Then four models are considered for the volume planning problem.

The objective of the ﬁrst model is to minimize the diﬀerences between

the total time of the preparation of all crew members, the objective of

the second one is to minimize the training expenses with a limitation of

the training level, and the objective of the third one is to maximize the

training level with a limited budget. The fourth model considers the

problem as an 𝑛-partition problem. Then two models are considered

for the calendar planning problem.

We consider the problem of planning the ISS cosmonaut training

with diﬀerent objectives. A pre-deﬁned set of minimum qualiﬁcation

levels should be distributed between the crew members with minimum

training time diﬀerences, training expenses or a maximum of the train-

ing level with a limitation of the budget.

First, a description of the cosmonaut training process is given.

Then four models are considered for the volume planning problem.

The objective of the ﬁrst model is to minimize the diﬀerences between

the total time of the preparation of all crew members, the objective of

the second one is to minimize the training expenses with a limitation of

the training level, and the objective of the third one is to maximize the

training level with a limited budget. The fourth model considers the

problem as an 𝑛-partition problem. Then two models are considered

for the calendar planning problem

Abstract

We consider the problem of planning the ISS cosmonaut training

with diﬀerent objectives. A pre-deﬁned set of minimum qualiﬁcation

levels should be distributed between the crew members with minimum

training time diﬀerences, training expenses or a maximum of the train-

ing level with a limitation of the budget.

First, a description of the cosmonaut training process is given.

Then four models are considered for the volume planning problem.

The objective of the ﬁrst model is to minimize the diﬀerences between

the total time of the preparation of all crew members, the objective of

the second one is to minimize the training expenses with a limitation of

the training level, and the objective of the third one is to maximize the

training level with a limited budget. The fourth model considers the

problem as an 𝑛-partition problem. Then two models are considered

for the calendar planning problem

Abstract

We consider the problem of planning the ISS cosmonaut training

with diﬀerent objectives. A pre-deﬁned set of minimum qualiﬁcation

levels should be distributed between the crew members with minimum

training time diﬀerences, training expenses or a maximum of the train-

ing level with a limitation of the budget.

First, a description of the cosmonaut training process is given.

Then four models are considered for the volume planning problem.

The objective of the ﬁrst model is to minimize the diﬀerences between

the total time of the preparation of all crew members, the objective of

the second one is to minimize the training expenses with a limitation of

the training level, and the objective of the third one is to maximize the

training level with a limited budget. The fourth model considers the

problem as an 𝑛-partition problem. Then two models are considered

for the calendar planning problem

Online reviews about the purchase of products or services provided have become the main source of users’ opinions. In order to gain profit or fame, usually spam reviews are written to promote or demote a few target products or services. This practice is known as review spamming. In the past few years, a variety of methods have been suggested in order to solve the issue of spam reviews. In this study, the researchers carry out a comprehensive review of existing studies on spam review detection using the Systematic Literature Review (SLR) approach. Overall, 76 existing studies are reviewed and analyzed. The researchers evaluated the studies based on how features are extracted from review datasets and different methods and techniques that are employed to solve the review spam detection problem. Moreover, this study analyzes different metrics that are used for the evaluation of the review spam detection methods. This literature review identified two major feature extraction techniques and two different approaches to review spam detection. In addition, this study has identified different performance metrics that are commonly used to evaluate the accuracy of the review spam detection models. Lastly, this work presents an overall discussion about different feature extraction approaches from review datasets, the proposed taxonomy of spam review detection approaches, evaluation measures, and publicly available review datasets. Research gaps and future directions in the domain of spam review detection are also presented. This research identified that success factors of any review spam detection method have interdependencies. The feature’s extraction depends upon the review dataset, and the accuracy of review spam detection methods is dependent upon the selection of the feature engineering approach. Therefore, for the successful implementation of the spam review detection model and to achieve better accuracy, these factors are required to be considered in accordance with each other. To the best of the researchers’ knowledge, this is the first comprehensive review of existing studies in the domain of spam review detection using SLR process.

We consider the problem of planning the ISS cosmonaut training

with diﬀerent objectives. A pre-deﬁned set of minimum qualiﬁcation

levels should be distributed between the crew members with minimum

training time diﬀerences, training expenses or a maximum of the train-

ing level with a limitation of the budget.

First, a description of the cosmonaut training process is given.

Then four models are considered for the volume planning problem.

The objective of the ﬁrst model is to minimize the diﬀerences between

the total time of the preparation of all crew members, the objective of

the second one is to minimize the training expenses with a limitation of

the training level, and the objective of the third one is to maximize the

training level with a limited budget. The fourth model considers the

problem as an 𝑛-partition problem. Then two models are considered

for the calendar planning problem.

For the volume planning problem, two algorithms are presented.

The ﬁrst one is a heuristic with a complexity of (𝑛)operations. The

second one consists of a heuristic and exact parts, and it is based on

the 𝑛-partition problem approach.

# 2) Feature engineering for mobile (SMS) spam filtering

**AUTHORS**: **Cormack G V.**

# Mobile spam in an increasing threat that may be addressed using filtering systems like those employed against email spam. We believe that email filtering techniques require some adaptation to reach good levels of performance on SMS spam, especially regarding message representation. In order to test this assumption, we have performed experiments on SMS filtering using top performing email spam filters on mobile spam messages using a suitable feature representation, with results supporting our hypothesis.

# 3) Feature engineering and tree modeling for author-paper identification challenge

**AUTHORS: Li J, Liang X, Ding W**

The ability to search literature and collect/aggregate metrics around publications is a central tool for modern research. Both academic and industry researchers across hundreds of scientific disciplines, from astronomy to zoology, increasingly rely on search to understand what has been published and by whom. Microsoft Academic Search is an open platform, which provides a variety of metrics and experiences for the research community, in addition to literature search. As the covering data came from many sources, the profile of an author with an ambiguous name tends to contain noise, resulting in papers that are incorrectly assigned to others. KDD Cup 2013 Track 1 challenges participants to determine which papers in an author profile were truly written by the given author. In this work, we present how to use tree-base models to accurately predict the paper author. We incorporate feature engineering into the models with the advantages of them. This paper introduces two kinds of tree-base models (GB-DT , RGF ) and presents in detail the learning algorithm and how features can be generated for the task. The experimental results show the effectiveness of the proposed approach.

**4)** **Feature Engineering for Depression Detection in Social Media**

**AUTHORS:** **Stankevich M, Isakov V, Devyatkin D**

This research is based on the CLEF/eRisk 2017 pilot task which is focused on early risk detection of depression. The CLEF/eRsik 2017 dataset consists of text examples collected from messages of 887 Reddit users. The main idea of the task is to classify users into two groups: risk case of depression and non-risk case. This paper considers different feature sets for depression detection task among Reddit users by text messages processing. We examine our bag-of-words, embedding and bigram models using the CLEF/eRisk 2017 dataset and evaluate the applicability of stylometric and morphological features. We also perform a comparison of our results with the CLEF/eRisk 2017 task report.

**5**) **AN ENHANCED AD EVENT-PREDICTION METHOD BASED ON FEATURE ENGINEERING**

**AUTHORS:** **Chen J H, Li X Y, Zhao Z Q**

In digital advertising, Click-Through Rate (CTR) and Conversion Rate (CVR) are very important metrics for evaluating ad performance. As a result, ad event prediction systems are vital and widely used for sponsored search and display advertising as well as Real-Time Bidding (RTB). In this work, we introduce an enhanced method for ad event prediction (i.e. clicks, conversions) by proposing a new efficient feature engineering approach. A large realworld event-based dataset of a running marketing campaign is used to evaluate the efficiency of the proposed prediction algorithm. The results illustrate the benefits of the proposed ad event prediction approach, which significantly outperforms the alternative ones.