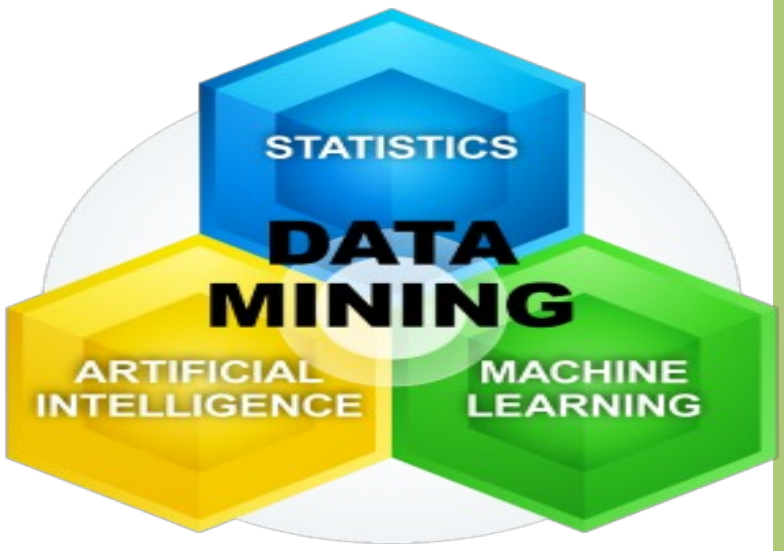
Technofist



DATA MINING PROJECT LIST 2018 -2019

At work as usual: 080-40969981 | Write to me: technofist.projects@gmail.com, | when u need us the most: **+91-9008001602, 080-40969981**| On the Web:www.technofist.comwww.technofist.in, www.itcdp.in

TECHNOFIST a leading student's project solution providing company established in Bangalore since 2007. With perfect infrastructure, lab set up, Work shop, Expertise faculties make us competitive service providers.

Here is the list of project titles 2018 and 2019.

DOORS OF TECHNOLOGY:

- EMBEDDED SYSTEMS
- MICROCONTROLLERS / ARM /PIC / AVR
- WIRELESS TECHNOLOGIES
- ROBOTICS
- ARDUINO
- GSM & GPS/ ZIGBEE
- MATLAB / VLSI
- IEEE PROJECTS ON JAVA / DOT NET
- INTERNET OF THINGS
- ANDROID BASED PROJECTS
- PHP
- AND COMPLETE MECHANICAL FABRICATIONS
- MECHANICAL DESIGN AND ANALYSIS

Projects are available for all branches of **ENGINEERING, DIPLOMA, MCA/BCA, and MSc/BSc.**

CORPORATEOFFICE:

At work as usual: 080-40969981 | Write to me: technofist.projects@gmail.com, | when u need us the most: **+91-9008001602, 080-40969981**| On the Web:www.technofist.comwww.technofist.in, www.itcdp.in

TECHNOFIST
YES Complex,
#19/3&4, 2nd Floor,
Dinnur Main Road, R T Nagar Post
Opposite to NavodayaVidyaneketan School
Bangalore – 32

Support : 080-40969981 Sales :+91-9008001602
www.technofist.com www.technofist.in www.itcdp.in
technofist.projects@gmail.com

Here we provided a ***latest Data mining 2018 project lis***t with abstracts. We do train a student from basic level of software which includes basic java Classes, projects implementation, final project demo and final code explanations. If you have questions regarding these projects feel free to contact us. You may also ask for abstract of a project idea that you have or want to work on.The ***own projects idea*** for diploma and Engineering students can also encouraged here.

IEEE DATA MINING PROJECT LIST 2018 AND 2019

	2018 – 19 IEEE PROJECT TITLES ON DATA MINING
TED001	**TITLE:**NETSPAM A NETWORK-BASED SPAM DETECTION FRAMEWORK FOR REVIEWS IN ONLINE SOCIAL MEDIA. **ABSTRACT**–Nowadays, a big part of people rely on available content in social media in their decisions (e.g., reviews and feedback on a topic or product). The possibility that anybody can leave a review provides a golden opportunity for spammers to write spam reviews about products and services for different interests. Identifying these spammers and the spam content is a hot topic of research, and although a considerable number of studies have been done recently toward this end, but so far

	the methodologies put forth still barely detect spam reviews, and none of them show the importance of each extracted feature type.
TED002	**TITLE** -POINT-OF-INTEREST RECOMMENDATION FOR LOCATION PROMOTION IN LOCATIONBASED SOCIAL NETWORKS **ABSTRACT** - Data access control is a challenging issue in public cloud storage systems. Ciphertext-Policy Attribute-Based Encryption (CP-ABE) has been adopted as a promising technique to provide flexible, fine-grained and secure data access control for cloud storage with honest-but-curious cloud servers. However, in the existing CP-ABE schemes, the single attribute authority must execute the time-consuming user legitimacy verification and secret key distribution, and hence it results in a single-point performance bottleneck when a CP-ABE scheme is adopted in a large-scale cloud storage system.
TED003	**TITLE - SOCIALQ&A: AN ONLINE SOCIAL NETWORK BASED QUESTION AND ANSWER SYSTEM** **ABSTRACT–** Question and Answer (Q&A) systems play a vital role in our daily life for information and knowledge sharing. Users post questions and pick questions to answer in the system. Due to the rapidly growing user population and the number of questions, it is unlikely for a user to stumble upon a question by chance that (s)he can answer. Also, altruism does not encourage all users to provide answers, not to mention high quality answers with a short answer wait time. The primary objective of this paper is to improve the performance of Q&A systems by actively forwarding questions to users who are capable and willing to answer the questions. To this end, we have designed and implemented SocialQ&A, an online social network based Q&A system. SocialQ&A leverages the social network properties of common-interest and mutual-trust friend relationship to identify an asker through friendship who are most likely to answer the question, and enhance the user security.

TED004	**TITLE -MODELING URBAN BEHAVIOR BY MINING GEOTAGGED SOCIAL DATA** **ABSTRACT** - Data generated on location-based social networks provide rich information on the whereabouts of urban dwellers. Specifically, such data reveal who spends time where, when, and on what type of activity (e.g., shopping at a mall, or dining at a restaurant). That information can, in turn, be used to describe city regions in terms of activity that takes place therein. For example, the data might reveal that citizens visit one region mainly for shopping in the morning, while another for dining in the evening. Furthermore, once such a description is available, one can ask more elaborate questions.
TED005	**TITLE - A WORKFLOW MANAGEMENT SYSTEM FOR SCALABLE DATA MINING ON CLOUDS** **ABSTRACT** - The extraction of useful information from data is often a complex process that can be conveniently modeled as a data analysis workflow. When very large data sets must be analyzed and/or complex data mining algorithms must be executed, data analysis workflows may take very long times to complete their execution. Therefore, efficient systems are required for the scalable execution of data analysis workflows, by exploiting the computing services of the Cloud platforms where data is increasingly being stored. The objective of the paper is to demonstrate how Cloud software technologies can be integrated to implement an effective environment for designing and executing scalable data analysis workflows. We describe the design and implementation of the Data Mining Cloud Framework (DMCF), a data analysis system that integrates a visual workflow language and a parallel runtime with the Software-as-a-Service (SaaS) model. DMCF was designed taking into account the needs of real data mining applications.
TED006	**TITLE -FIDOOP: PARALLEL MINING OF FREQUENT ITEMSETS USING MAPREDUCE.** **ABSTRACT** -Existing parallel mining algorithms for frequent itemsets lack a mechanism that enables automatic parallelization, load balancing, data distribution, and fault tolerance on large clusters. As a solution to this problem, we design a parallel frequent itemsets mining algorithm called FiDoop using the MapReduce

	programming model. To achieve compressed storage and avoid building conditional pattern bases, FiDoop incorporates the frequent items ultra-metric tree, rather than conventional FP trees.
TED007	**TITLE - SENTIMENT ANALYSIS OF TOP COLLEGES IN INDIA USING TWITTER DATA** **ABSTRACT** -In today's world, opinions and reviews accessible to us are one of the most critical factors in formulating our views and influencing the success of a brand, product or service. With the advent and growth of social media in the world, stakeholders often take to expressing their opinions on popular social media, namely Twitter. While Twitter data is extremely informative, it presents a challenge for analysis because of its humongous and disorganized nature. This paper is a thorough effort to dive into the novel domain of performing sentiment analysis of people's opinions regarding top colleges in India. Besides taking additional preprocessing measures like the expansion of net lingo and removal of duplicate tweets, a probabilistic model based on Bayes' theorem was used for spelling correction, which is overlooked in other research studies.
TED008	**TITLE - INVERTED LINEAR QUADTREE: EFFICIENT TOP K SPATIAL KEYWORD SEARCH** **ABSTRACT** -With advances in geo-positioning technologies and geo-location services, there are a rapidly growing amount of spatio-textual objects collected in many applications such as location based services and social networks, in which an object is described by its spatial location and a set of keywords (terms). Consequently, the study of spatial keyword search which explores both location and textual description of the objects has attracted great attention from the commercial organizations and research communities. In the paper, we study two fundamental problems in the spatial keyword queries: top k spatial keyword search (TOPK-SK), and batch top k spatial keyword search (BTOPK-SK). Given a set of spatio-textual objects, a query location and a set of query keywords, the TOPK-SK retrieves the closest k objects each of which contains all keywords in the query. BTOPK-SK is the batch processing of sets of TOPK-SK queries.
TED009	**TITLE - TRUTH DISCOVERY IN CROWDSOURCED DETECTION OF SPATIAL EVENTS**

	ABSTRACT -The ubiquity of smartphones has led to the emergence of mobile crowdsourcing tasks such as the detection of spatial events when smartphone users move around in their daily lives. However, the credibility of those detected events can be negatively impacted by unreliable participants with low-quality data. Consequently, a major challenge in mobile crowdsourcing is truth discovery, i.e., to discover true events from diverse and noisy participants' reports. This problem is uniquely distinct from its online counterpart in that it involves uncertainties in both participants' mobility and reliability.
TED010	**TITLE - SPORE: A SEQUENTIAL PERSONALIZED SPATIAL ITEM RECOMMENDER SYSTEM** **ABSTRACT** -With the rapid development of location-based social networks (LBSNs), spatial item recommendation has become an important way of helping users discover interesting locations to increase their engagement with location-based services. Although human movement exhibits sequential patterns in LBSNs, most current studies on spatial item recommendations do not consider the sequential influence of locations. Leveraging sequential patterns in spatial item recommendation is, however, very challenging.
TED011	**TITLE - FRAPPE: DETECTING MALICIOUS FACEBOOK APPLICATIONS** **ABSTRACT** -Communication technology has completely occupied all the areas of applications. Last decade has however witnessed a drastic evolution in information and communication technology due to the introduction of social media network. Business growth is further achieved via these social media. Nevertheless, increase in the usage of online social networks (OSN) such as Facebook, twitter, Instagram etc has however led to the increase in privacy and security concerns. Third party applications are one of the many reasons for Facebook attractiveness. Regrettably, the users are unaware of detail that a lot of malicious Facebook applications provide on their profile.
TED012	**TITLE -A NOVEL RECOMMENDATION MODEL REGULARIZED WITH USER TRUST AND ITEM RATINGS** **ABSTRACT** -We propose TrustSVD, a trust-based matrix factorization technique for recommendations. TrustSVD integrates multiple information sources into the recommendation model in order to reduce the data sparsity and cold start problems and

	their degradation of recommendation performance. An analysis of social trust data from four real-world data sets suggests that not only the explicit but also the implicit influence of both ratings and trust should be taken into consideration in a recommendation model.
TED013	**TITLE - AUTOMATICALLY MINING FACETS FOR QUERIES FROM THEIR SEARCH RESULTS** **ABSTRACT-** We address the problem of finding query facets which are multiple groups of words or phrases that explain and summarize the content covered by a query. We assume that the important aspects of a query are usually presented and repeated in the query's top retrieved documents in the style of lists, and query facets can be mined out by aggregating these significant lists. We propose a systematic solution, which we refer to as QDMiner, to automatically mine query facets by extracting and grouping frequent lists from free text, HTML tags, and repeat regions within top search results. Experimental results show that a large number of lists do exist and useful query facets can be mined by QDMiner.
TED014	**TITLE - BUILDING AN INTRUSION DETECTION SYSTEM USING A FILTER-BASED FEATURE SELECTION ALGORITHM** Redundant and irrelevant features in data have caused a long-term problem in network traffic classification. These features not only slow down the process of classification but also prevent a classifier from making accurate decisions, especially when coping with big data. In this paper, we propose a mutual information based algorithm that analytically selects the optimal feature for classification. This mutual information based feature selection algorithm can handle linearly and nonlinearly dependent data features. Its effectiveness is evaluated in the cases of network intrusion detection.
TED015	**TITLE - CONNECTING SOCIAL MEDIA TO E-COMMERCE: COLD-START PRODUCT RECOMMENDATION USING MICROBLOGGING INFORMATION** **ABSTRACT -** Unsupervised Cross-domain Sentiment Classification is the task of adapting a sentiment classifier trained on a particular domain (source domain), to a different domain (target domain), without requiring any labeled data for the target domain. By adapting an existing sentiment classifier to previously unseen target domains, we can avoid the cost for manual data annotation for the target domain. We model this problem as embedding learning, and construct three objective functions that capture: (a) distributional properties of pivots (i.e., common features

	that appear in both source and target domains), (b) label constraints in the source domain documents, and source and target domains.

HEAD OFFICE

TECHNOFIST- R T NAGAR

YES Complex, #19/3&4 , 2nd Floor, Dinnur Main Road , **R. T. Nagar** Post , Bangalore 560032
Ph : 080 40969981, Mob : +91 9008001602
www.technofist.com, www.technofist.in, www.itcdp.in