

## SISTEM ZA NEPRISTRASNU ANALIZU KAŠNJENJA BAZIRAN NA TEHNIKAMA VEŠTAČKE INTELIGENCIJE

### A FRAMEWORK FOR UNBIASED DELAY ANALYSIS BASED ON ARTIFICIAL INTELLIGENCE TECHNIQUES

Marija Ivanović<sup>1</sup>, Đorđe Nedeljković<sup>1</sup>, Zoran Stojadinović<sup>1</sup>

<sup>1</sup>Građevinski fakultet, Univerzitet u Beogradu, Srbija

**Apstrakt:** Kašnjenja su uobičajena pojava na građevinskim projektima širom sveta. Identifikacija uzroka kašnjenja je prvi i najznačajniji korak u procesu eliminisanja ili umanjenja kašnjenja na budućim građevinskim projektima. Cilj ovog istraživanja jeste predlog sistema zanepristrasnu detekciju uzroka kašnjenja na građevinskim projektima na bazi tehnika veštačke inteligencije. Postojeća istraživanja su bazirana su ekspertskom mišljenju koje je obojeno subjektivizmom, pristrasnošću i bez iskorišćenog potencijala tekstualne dokumentacije sa projekta. Post-projekatne analize su koristan izvor informacija o kašnjenjima, međutim, kompleksnost koordinacije projektnog tima završenog projekta doprinosi negativnom trendu izrade ovih dokumenata. Kroz istraživanje su predstavljeni ključni elementi sistema zadetektiranja uzroka kašnjenja baziranog na Transformer jezičkim modelima - DREAM (*Delay Root causes Extraction and Analysis Model*). Predloženi DREAM kombinuje stručno znanje, Transformer jezičke modele i zapisnike sa nedeljnih sastanaka (MoM) kao značajan korpus nestrukturirane tekstualne dokumentacije. Rezultati predloženog sistema treba da omoguće ekspertima da razumeju prirodu uzroka kašnjenja kako bi se pokrenuo proces inteligentnog donošenja odluka na budućim projektima.

**Ključne reči:** *Kašnjenja, Uzroci Kašnjenja, Analiza Teksta, Transformeri, Veštačka Inteligencija, Upravljanje Projektima*

**Abstract:** Time overruns are a common occurrence in construction projects around the world. Identifying the causes of delays is the first and most important step in eliminating or mitigating delays on future construction projects. This research aims to propose a framework for the detection of the causes of delays in construction projects based on artificial intelligence techniques. Existing research is based on expert opinion, burdened by subjectivism and bias, and without using the potential of the text documentation from the project. Post-project analyses are a valuable source of information about delays. However, the complexity of coordinating the project team of the completed project contributes to the negative trend of producing these documents. The research presents the key elements of the

delay detection system based on Transformer language models - DREAM (*Delay Root causes Extraction and Analysis Model*). The proposed DREAM combines expert knowledge, *Transformer* language models, and Minutes of weekly Meetings (MoMs) as substantial unstructured textual documentation. The results of the proposed system should enable experts to understand the nature of the causes of delay and initiate the process of intelligent decision-making on future projects

**Keywords:** *Delay, Causes Of Delay, Text Mining, Transformers, Artificial Intelligence (AI), Construction Management*

## 1. UVOD

Prekoračenja na građevinskim projektima predstavljaju fenomen prisutan na projektima širom sveta. Najčešće se razmatraju prekoračenja u kontekstu vremena i troškova, koji pored kvaliteta čine elemente "gvozdene" trougla (*iron triangle*) (Pollack et al., 2018). Većina autora smatra da je rok za završetak projekta jedan od najznačajnijih ciljeva projekta (Wang et al., 2018). Istorija građevinske industrije beleži veliki broj projekata koji su prekoračili planirano vreme izgradnje (Mahamid, 2012), iako se ova oblast istražuje decenijama unazad.

Proces identifikacije uzroka kašnjenja predstavlja prvu i najznačajniju fazu ka formiranju pristupa za kontrolu i umanjenje štetnih posledica kašnjenja po projekat (Chidambaram, 2012). Istraživanje otvorene literature ukazuje da se metodologija za identifikaciju uzroka kašnjenja bazira na dva izvora. Prvi izvor se odnosi na empirijska istraživanja i prikupljanje ekspertskih iskustava o uzrocima kašnjenja za različite vrste građevinskih projekata. Takav pristup daje rezultate koji su neminovno obojeni subjektivizmom i ekspertskom pristrasnošću. Drugi izvor se bazira na istraživanjima zasnovanim na podacima. Pod podacima se podrazumevaju različiti oblici numeričkih podataka. Takav pristup podrazumeva značajne količine resursa za obradu, analizu i donošenje zaključaka.

Konvencionalni pristupi identifikaciji uzroka kašnjenja, iako značajni, imaju niz nedostataka (Flyvbjerg, 2021). Primarno, subjektivizam i pristrasnost eksperata utiču na pouzdanost rezultata o uzrocima kašnjenja. Sekundarno, postojeća istraživanja nisu iskoristila potencijal tekstualne dokumentacije sa projekta. Prema nekim istraživanjima, građevinski projekti baštine do 10.000 različitih tekstualnih dokumenata (Ivanović et al., 2022) ili preko 6 terabajta različitih strukturiranih ili nestrukturiranih podataka (Nedeljković, 2018).

Pomenuta ograničenja u otvorenoj literaturi moguće je prevazići primenom savremenih tehnika, naročito u domenu veštačke inteligencije (*Artificial Intelligence - AI*). Cilj ove studije nije eliminisanje ekspertskog znanja, već njegova racionalna upotreba. Zahvaljujući veštačkoj inteligenciji moguće je transformisati veliku količinu dokumentacije u korisno znanje, čime se unapređuje proces donošenja odluka. Upravljanje projektom bazirano na podacima povećava produktivnost rada za oko 5% (Abbaszadegan & Grau, 2015).

Cilj ovog istraživanja jeste predlog sistema za nepristrasnu identifikaciju i analizu uzroka kašnjenja baziranog na tehnikama analize teksta (*text mining*). U domenu ovog rada biće

prikazan okvir za razvoj sistema za identifikaciju i analizu uzroka kašnjenja baziranog na tehnikama veštačke inteligencije DREAM (Delay Root causes of delay Extraction and AnalysisModel).

## 2. PREGLED LITERATURE

Pregled otvorene literature je podrazumevao istraživanje studija o uzrocima kašnjenja na građevinskim projektima, kao i istraživanje o metodama za analizu teksta (*text mining*) i njihovoj primeni u upravljanju projektima.

Tokom prve faze pregleda literature, akcenat je bio na identifikovanim uzrocima u cilju formiranja jedinstvene liste uzroka kašnjenja. Kako bi se formirala sveobuhvatna lista uzroka kašnjenja, pregledano je 80 studija. Rezultat ovog dela istraživanja jeste lista od 54 uzroka kašnjenja grupisanih u 8 kategorija (Projektna dokumentacija, Tenderski postupak, Resursi, Izvođač, Investitor, Stručni nadzor, Projekat, Spoljašnji). Formiranje liste uzroka kašnjenja predstavlja prvi korak u razvoju modela (Slika 1). Na taj način, svaki uzrok kašnjenja je dobio svoj kod, kao na primer: Uzrok kašnjenja – C1Nedostatak radne snage na gradilištu; Grupa G3Resursi = Kod G3C1. Detaljni prikaz jedinstvene liste uzroka kašnjenja je deo istraživanja (Ivanović et al., 2022).

Druge deo pregleda literature se odnosio na istraživanje različitih jezičkih modela i njihove upotrebe u upravljanju projektima. U dosadašnjoj literaturi, prepoznata su tri domena primene analize teksta u upravljanju projektima: (1) pronalaženje informacija (Zou et al., 2017), (2) klasifikacija tekstualne dokumentacije (Caldas & Soibelman, 2003) i (3) sticanje znanja (Zhong et al., 2020). Usled izraženog korpusa nestrukturiranih tekstualnih podataka generisanih tokom životnog ciklusa projekta, postoji značajan potencijal za primenu rudarenja teksta u upravljanju građevinskim projektima. U nastavku sledi prikaz odabranih studija od značaja za predmet istraživanja.

Williams & Gong, (2014) su koristili tehnike rudarenja teksta (ANN i KNN) za razvoj modela za predikciju prekoračenja troškova projekta u fazi formiranja ponude. Pored numeričkih podataka o troškovima i broju ponuđača, model je razmatrao i tekstualne podatke. Kako izvor nestrukturirane tekstualne forme korišćeni su kratki opisi, od dve do tri rečenice, iz tenderske dokumentacije. Zaključak ove studije je doprineo pretpostavci da model baziran na kombinaciji tekstualnih i numeričkih podataka ima bolje performanse od modela u koji su integrisani samonumerički parametri.

Marzouk & Enaba, (2019) su primenili tehnike rudarenje po tekstualnim podacima za analizu i vizuelni prikaz građevinskih ugovora i korespondencije na projektu. Vizuelna prezentacija različitih informacija iz ugovorne dokumentacije je omogućila sagledavanje skrivenih formi i trendova. Autori su istakli važnost vizualnog predstavljanja tekstualne dokumentacije za pronalaženje skrivenih obrazaca u obimnom korpusu nestrukturiranog teksta.

Zhang et al., (2019) su primenili tehnike prirodne obrade jezika (NLP) i rudarenja po

tekstualnim podacima za analizu izveštaja o nezgodama na gradilištu (*accident claims*). S druge strane, Li et al., (2020) su predložili model za brzu klasifikaciju izveštaja o nesrećama na gradilištu. Zaključci pomenutih istraživanja su da su glavne prednosti integrisanja tehnika AI u domenu klasifikacije, značajno smanjenje vremena za analizu i obradu tekstualne dokumentacije čime se postiže optimalno upravljanje resursima.

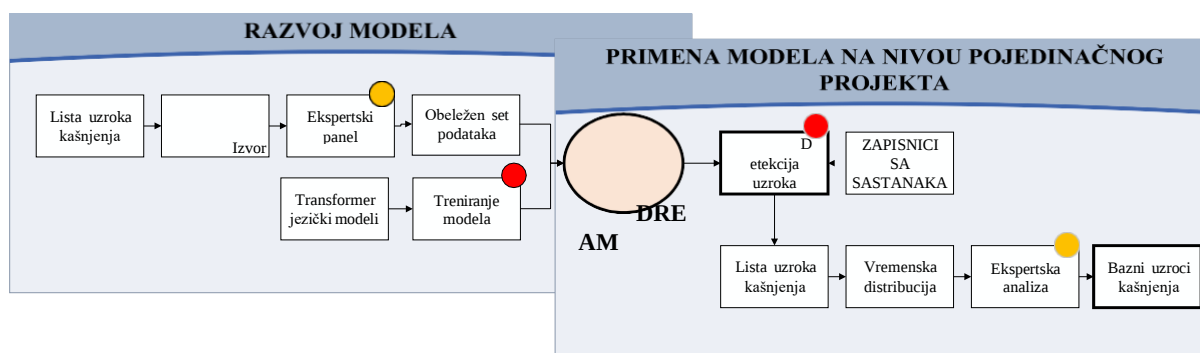
Nema sumnje da su postojeća istraživanja unapredila procese upravljanja građevinskim projektima. Iako veštačka inteligencija ima široku primenu, primetan je nedostatak istraživanjakoje se odnose na rizike i analizu kašnjenja na građevinskim projektima. Cilj ovog istraživanja jeste da se popuni uočena praznina u otvorenoj literaturi.

### 3. SISTEM ZA NEPRISTRASNU DETEKCIJU UZROKA KAŠNJENJA

Kroz ovo istraživanje se predlaže okvir za razvoj sistema za identifikaciju i analizu uzroka kašnjenja na građevinskim projektima primenom tehnika veštačke inteligencije, preciznije tehnika analize teksta. Rezultati istraživanja treba da doprinesu iskorišćenju obimne tekstualnedokumentacije u cilj nepristrasne detekcije uzroka kašnjenja na građevinskim projektima.

Sveobuhvatni sistem DREAM se sastoji od dve faze (slika1): I – Razvoj modela i II – Primena modela. U fazi razvoja modela razmatraju se glavne komponente i njihovi međusobni odnosima, dok se u fazi primene razmatraju ulazni podaci kojima će model biti pohranjen kao i krajnji rezultati.

Glavne komponente predloženog sistema se odnose se na: (1) zapisnike sa nedeljnih sastanaka,kao ulazna dokumentacija na bazi koje model vrši detekciju uzroka kašnjenja, (2) ekspertske obeležene tekst, kao izvor znanja za dodatno treniranje modela u domenu kašnjenja na građevinskim projektima i (3) *Transformer* jezičke modele za kreiranje i razvoj predloženog sistema.

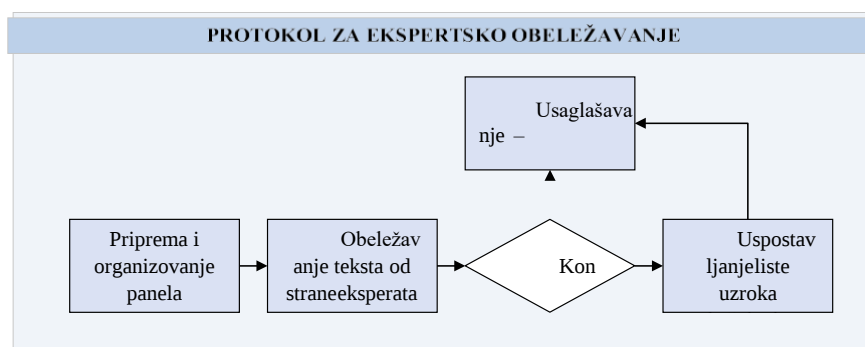


Slika 1. Struktura sistema i ključni rezultati

U nastavku sledi opis ključnih elemenata predloženog sistema.

### 3.1 EKSPERTSKO OBELEŽAVANJE DELOVA TEKSTA

Za potrebe razvoja modela za detekciju uzroka kašnjenja iz nestruktuiranih tekstualnih dokumenata, izvršeno je ekspertsko obeležavanje delova teksta koji se mogu dovesti u vezu sauzrocima kašnjenja. Za ovu fazu istraživanja organizovan je panel eksperata koji su činili 3 ugledna stručnjaka u oblasti upravljanja projektima. Panel je sproveden kroz 4 faze (slika 2). Od eksperata je zahtevano da svaki tekstualni segment obeleže sa kojim uzrokom kašnjenja se može dovesti u vezu (u skladu sa listom uzroka kašnjenja). Ekspertski konsenzus je postignut ukoliko je dva ili više eksperata identično povezalo tekst sa uzrokom kašnjenja. Ukoliko nije postojao konsenzus, sprovedena je usmena diskusija kako bi se usaglasila mišljenja.



**Slika 2:** Protokol za ekspertsko obeležavanje tekstualne dokumentacije

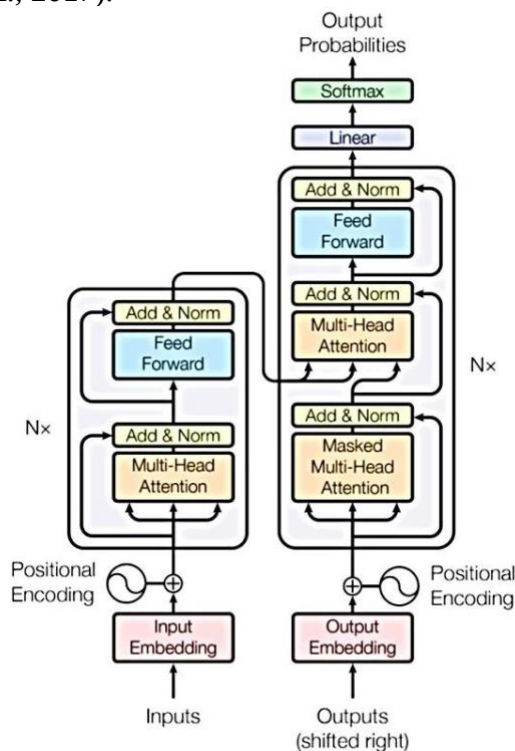
Cilj ekspertskog obeležavanja jeste dodatno obučavanje izabranih jezičkih modela kako bi se dodao još jedan semantički sloj sa krajnjim ciljem da se model prilagodi, do sada, nepoznatom okruženju građevinskih projekata. Važno je napomenuti da se ekspertsko obeležavanje sprovodi samo jednom, u fazi razvoja i treniranja modela.

### 3.2 TRANSFORMER JEZIČKI MODELI

Transformer jezički modeli su bazirani na *Transfer Learning* metodologijama, koje su omogućile napredak u prirodnom procesiranju jezika (*Natural Language Processing - NLP*) i analizi tekstualnih podataka (*text mining*). Ovakav pristup se bazira na upotrebi unapred obučanih modela i njihovog finog podešavanje kako bi se postigle visoke performanse u domenu rudarenja po tekstualnim podacima. Za razliku od ostalih metoda mašinskog učenja, *Transformer* jezički modeli koriste mehanizme samopažnje (*self-attention*). Prema

Vaswani et al., (2017), ovakvi jezički modeli dubokog mašinskog učenja su pogodni za razumevanja dužih sekvenci teksta, što je naročito pogodno za domen građevinskih projekata.

Na slicu 3 je prikazana arhitektura Transformer-a koja se obično sastoji od enkodera i dekodera (Vaswani et al., 2017).



**Slika 3:** Arhitektura *Transformer* modela (Vaswani et al., 2017)

BERT jezički model je sastavljen od blokova enkodera transformera i obučen je na Vikipediji (oko 2.500 miliona reči). Javno je dostupan od 2018. godine.

### 3.3 ZAPISINICI SA SASTANAKA

Izbor dokumentacije predstavlja ključni korak u razvoju sistema koji se baziraju na rudarenju po tekstualnim dokumentima. Za potrebe razvoja DREAM-a, uspostavljen je spisak kriterijumaza izbor relevantne dokumentacije:

- K1: Dostupnost (poverljivost sadržaja tekstualne dokumentacije);
- K2: Zastupljenost (univerzalnost dokumenta na svim građevinskim projektima);
- K3: Dostupnost u elektronskom formatu (moguće je samo razmatrati el. format);
- K4: Interaktivnost (involviranost ključnih učesnika u kreiranje dokumenta);
- K5: Vremenska komponenta (dokument sačinjen u vreme nastanka poremećaja);

- K6: Kontekst poremećaja (sadržaj dokumentacije mora biti u kontekstu kašnjenja).

Tokom istraživanja razmatrani su različiti tipovi tekstualne dokumentacije kao što su građevinski dnevници, mesečni izveštaji, odštetni zahtevi, post-projektni izveštaji, korespondencija i zapisnici sa nedeljnih sastanaka. Prema navedenim kriterijumima, samo su zapisnici sa nedeljnih sastanka ispunjavali sve prethodno navedene kriterijume.

Zapisnici sa nedeljnih sastanaka (*Minutes of Meetings* - MoM) predstavljaju dokumentaciju u čije kreiranje su uključeni svi učesnici sastanka. Takođe, postoji dobra praksa prihvatanja zapisnika od strane svih učesnika na sastanku čime se smanjuje mogućnost manipulisanja sadržajem. Kako se sastanci održavaju na nedeljnom nivou, može se zaključiti da sadržaj pravovremeno opisuje događaje na projektu. I na kraju, jedna od najznačajnijih prednosti zapisnika sa sastanaka jeste veza između potencijalnog uzroka i vremenske komponente, čime se usvaja pretpostavka da zapisnici sa sastanaka agregiraju sisteme ranog upozorenja po kašnjenja na projektu.

#### 4. ZAKLJUČAK

Značaj iskustava i znanja sa tekućih projekata se često zanemaruje što je direktno u korelaciji sa negativnim trendom izrade post-projektnih izveštaja. U završnoj fazi, scenariji sa različitim projekata su neretko slični. Nakon završetka projekta projekti timovi se raspuštaju i fokus se prebacuje na buduće projekte. Pojedinci postaju bogatiji za prethodna iskustva, ali nedostaje zaokruženo, na bazi kolektivnog konsenzusa, znanje ključnih učesnika šta je „*pošlo po zlu*“ na prethodnom projektu. Otkrivanje uzroka kašnjenja završenih projekata je prvi korak ka eliminisanju ili ublažavanju potencijalnih kašnjenja na budućim projektima.

Cilj ovog istraživanja je bio predlog okvira za nepristrasno otkrivanje uzroka kašnjenja na bazi tehnika veštačke inteligencije, odnosno tehnika analize teksta. Dodatno, krajnji cilj predloženog sistema jeste da se omogući okruženje za otkrivanje baznih uzroka kašnjenja. Pomenuti ciljevisu prepoznati kao značajni od strane drugih autora (Flyvbjerg Bent, 2011) ali, na bazi opsežnog pregleda otvorene literature od strane autora studije, nedovoljno istraženi.

Kroz ovo istraživanje je predstavljen jedinstven pristup nepristrasnoj detekciji uzroka kašnjenja upotrebom tehnika analize teksta. Ključne komponente predloženog DREAM su:

- Zapisnici sa sastanaka
- Ekspertski obeleženi tekstualni segmenata
- *Transformer* jezički modeli

Predloženi sistem DREAM uspostavlja balans između mašinskog učenja i ekspertskog angažovanja. Sinergija komponenata omogućava racionalno korišćenje resursa u odabranim domenima. Rezultati modela mogu biti od koristi donosiocima odluka na budućim

projektima da, u kratkom roku, dobiju rekonstrukciju događaja koje se odnose na kašnjenja na nivou pojedinačnog projekta. Na taj način moguće je efikasno i fokusirano upravljati vremenom na tekućem projektu dokazujući da veštačka inteligencija omogućava superiorniji sistem u odnosu na konvencionalne pristupe.

Predloženi sistem poseduje i određena ograničenja koja usmeravaju buduća istraživanja. DREAM koristi ekspertsko znanje kao ulazni podatak za treniranje modela, stoga takav segment predstavlja potencijalni rizik od uvođenja subjektivizma i pristrasnosti u model. Nadalje, ukoliko su ulazni podaci za fino podešavanje modela neuravnoteženi ili irelevantni, sistem će imati smanjene performanse za otkrivanje uzroka kašnjenja (prema usvojenoj klasifikaciji). U cilju izbegavanja pomenutih ograničenja, pravci budućih istraživanja će se odnositi na smernice i formalni protokol za izbor relevantnih izvora teksta za treniranje kao i protokola za njihovo obeležavanje.

## LITERATURA

1. Abbaszadegan, A., & Grau, D. (2015). Assessing the Influence of Automated Data Analytics on Cost and Schedule Performance. *Procedia Engineering*, 123, 3–6. <https://doi.org/10.1016/j.proeng.2015.10.047>
2. Caldas, C. H., & Soibelman, L. (2003). Automating hierarchical document classification for construction management information systems. *Automation in Construction*, 12(4), 395–406. [https://doi.org/10.1016/S0926-5805\(03\)00004-9](https://doi.org/10.1016/S0926-5805(03)00004-9)
3. Flyvbjerg, B. (2021). Top Ten Behavioral Biases in Project Management: An Overview. *Project Management Journal*, 52(6), 531–546. <https://doi.org/10.1177/87569728211049046>
4. Flyvbjerg Bent. (2011). *Over budget, over time, over and over again: Managing major projects*. Oxford University Press. <http://ssrn.com/abstract=2278226>
5. Ivanović, M. Z., Nedeljković, Đ., Stojadinović, Z., Marinković, D., Ivanišević, N., & Simić, N. (2022). Detection and In-Depth Analysis of Causes of Delay in Construction Projects: Synergy between Machine Learning and Expert Knowledge. *Sustainability*, 14(22). <https://doi.org/10.3390/su142214927>
6. Li, R. Y. M., Li, H. C. Y., Tang, B., & Au, W. C. (2020). Fast AI classification for analyzing construction accidents claims. *ACM International Conference Proceeding Series*, 1–4. <https://doi.org/10.1145/3407703.3407705>
7. Marzouk, M., & Enaba, M. (2019). Text analytics to analyze and monitor construction project contract and correspondence. *Automation in Construction*, 98, 265–274. <https://doi.org/10.1016/j.autcon.2018.11.018>
8. Nedeljković, Đ. L. (2018). *Издавање и визуелизација знања из текстуалних извора за потребе управљања инвестиционим пројектима у грађевинарству, Докторска дисертација*. Универзитет у Београду.
9. Pollack, J., Helm, J., & Adler, D. (2018). What is the Iron Triangle, and how has it changed? *International Journal of Managing Projects in Business*, 11(2), 527–547. <https://doi.org/10.1108/IJMPB-09-2017-0107>

10. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. *31st Conference on Neural Information Processing Systems (NIPS 2017)*. <http://arxiv.org/abs/1706.03762>
11. Wang, T. K., Ford, D. N., Chong, H. Y., & Zhang, W. (2018). Causes of delays in the construction phase of Chinese building projects. *Engineering, Construction and Architectural Management*, 25(11), 1534–1551. <https://doi.org/10.1108/ECAM-10-2016-0227>
12. Williams, T. P., & Gong, J. (2014). Predicting construction cost overruns using text mining, numerical data and ensemble classifiers. *Automation in Construction*, 43, 23–29. <https://doi.org/10.1016/j.autcon.2014.02.014>
13. Zhang, F., Fleyeh, H., Wang, X., & Lu, M. (2019). Construction site accident analysis using text mining and natural language processing techniques. *Automation in Construction*, 99, 238–248. <https://doi.org/10.1016/j.autcon.2018.12.016>
14. Zhong, B., Pan, X., Love, P. E. D., Sun, J., & Tao, C. (2020). Hazard analysis: A deep learning and text mining framework for accident prevention. *Advanced Engineering Informatics*, 46. <https://doi.org/10.1016/j.aei.2020.101152>
15. Zou, Y., Kiviniemi, A., & Jones, S. W. (2017). Retrieving similar cases for construction project risk management using Natural Language Processing techniques. *Automation in Construction*, 80, 66–76. <https://doi.org/10.1016/j.autcon.2017>.