



## Açık Kaynak İş Zekâsı Uygulaması

Dursun DİNÇER<sup>a1</sup>, Sefa SARI<sup>a2\*</sup>, Erdem ÇER<sup>a3</sup>, Mustafa KAMAŞAK<sup>b4</sup>

<sup>a</sup> Yurtiçi Kargo Servisi A.Ş. Ar-Ge Merkezi, İstanbul, Türkiye

<sup>b</sup> İstanbul Teknik Üniversitesi, Bilgisayar ve Bilişim Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul, Türkiye

Istanbul Sabahattin Zaim Üniversitesi Fen Bilimleri Enstitüsü Dergisi (2021) 3 (3): 189-192

<https://doi.org/10.47769/izufbed.872189>

ORCID <sup>1</sup> 0000-0001-6248-0394; <sup>2</sup> 0000-0003-1528-5187; <sup>3</sup> 0000-0002-9198-4660; <sup>4</sup> 0000-0002-5050-3357

### YAYIN BİLGİSİ

Yayın geçmişi:

Gönderilen tarih: 01 Şubat 2021

Kabul tarihi: 22 Eylül 2021

### Anahtar kelimeler:

İş Zekâsı

Yakın Gerçek Zamanlı Raporlama

Açık Kaynak

Büyük Veri

Veri Analizi

### ÖZET

Bu çalışmada açık kaynak kodlu teknolojiler ile geliştirilen yakın gerçek zamanlı iş zekâsı uygulaması açıklanmakta ve elde edilen sonuçlar irdelenmektedir. Uçtan uca kullanılacak bir raporlama aracı her sektörden şirketin ihtiyacı olup açık kaynak raporlama araçları lisans maliyetleri konusunda şirketlerin yükünü oldukça hafiflettiği için tercih edilmiştir. Yapılan çalışmalarla alt yapı olarak büyük veri üzerindeki çalışmaları, tahminleme modelleri, karar destek algoritmaları ve öğrenen modellerin geliştirilebileceği bir altyapı geliştirilerek, geniş bir veri tabanı portföyünü, yakın gerçek zamanlı raporlamayı, mobil cihazlarla uyumluluğu destekleyen iş zekâsı uygulaması yaklaşık 1000 kullanıcının kullanımına açılmıştır.

## Open Source Business Intelligence Application

### ARTICLE INFO

Article history:

Received: 01 February 2021

Accepted: 22 September 2021

### Key words:

Business Intelligence

Near Real-Time Reporting

Open Source

Big Data

Data Analysis

### ABSTRACT

In this work, near real time business intelligence application was developed using open source libraries, and its performance was investigated. An end-to-end reporting tool is needed by companies from every sector, and open source reporting tools are preferred due to high license costs of commercial products. An infrastructure was developed that works on big data upon which prediction models, decision support algorithms and learning models can be developed. The business intelligence application supports various databases, near-real time reporting and compatibility with mobile devices. Currently, the developed business intelligence application is used by approximately 1000 users.

### 1. Giriş

Firmaların karar mekanizmaları ve rasyonel yönetimi için firma içerisindeki tüm süreçlere ait iş zekâsı araçları ve raporlama hizmetleri kritik öneme sahiptir [1]. Firmalar bu ihtiyacı genellikle mevcut çözümler arasından seçim yaparak karşılamaktadır. Firmaların iş zekâsı araçları için hazır çözümleri analiz ederken veya çözümleri birbirleri ile mukayese ederken kullandıkları bir takım kriterler bulunmaktadır. Bu kriterlerden bazıları esneklik, teknik

destek ve maliyettir [2].

İş zekâsı araçlarında her firmanın ihtiyaçları genel manada birbirlerine benzemesine rağmen firmaların kendilerine özel bir takım farklı ihtiyaçları da ortaya çıkabilmektedir. Bu sebeple, iş zekâsı araçlarının hem arayüz hem de işlevsel manada özelleştirilebilir olması ve özelleştirme konusunda kullanıcılara esneklik tanınması da oldukça önemlidir. Bazı araçlar bir programlama dili üzerinden makro/betik (ing: script) geliştirilmesini sağlarken, diğer ticari çözümlerde bu

\*Sorumlu yazar.

E-mail adresi: [sefa.sari@yurticikargo.com](mailto:sefa.sari@yurticikargo.com) (Sefa SARI)

tür özelleştirmeler ek geliştirme kapsamında değerlendirilerek ilave maliyet oluşturmaktadır.

Firmanın operasyon türüne göre iş zekâsı araçlarının oluşturduğu raporların güncelliği bazı durumlarda oldukça önemli olabilmektedir [3]. Bazı iş alanlarında 1 gün öncesinin verisi ile hazırlanmış raporlar bile yeterli olurken, diğer iş alanlarında en güncel veriler ile hazırlanması oldukça önem arz edebilmektedir. Dolayısı ile iş zekâsı araçlarının seçiminde oluşturdukları raporlarda kullanılan verinin güncelliği de bir kistas olarak değerlendirilmektedir.

Firmalarda operasyonların güvenilirliği ve devamlılığı esas olduğu için operasyonel kararları destekleyen iş zekâsı çözümlerinin de sürekli erişilebilir ve devamlılık arz etmesi gerekmektedir [4]. Bu sebeple, ortaya çıkacak kesintilerin önceden bildirilmesi, önceden belirlenmiş bir seviyenin üzerine çıkmaması gibi hususlar bu araçlarının seçiminde rol oynamaktadır.

İş zekâsı araçlarının kullanımı sırasında ortaya çıkabilecek aksaklıkların ve eksikliklerin hızlı şekilde aşılması için teknik desteğin varlığı ve kalitesi karar süreçlerinde öne çıkmaktadır.

İş zekâsı araçları farklı lisanslama modelleri ile pazara sunulmaktadır. Bu modellerden bazıları kullanıcı başına lisanslama, kullanım başına lisanslama vb. şeklinde olabilmektedir. Farklı çok sayıda lisans modelinin firma için maliyetinin hesaplanması bile bazı durumlarda oldukça güç olabilmektedir. Sabit ücretin yanı sıra bu çözümlerin idame masraflarının da göz önünde bulundurulması gerekmektedir. Bazı çözümler için güncelleme ücreti, teknik destek hizmeti vb. isimler altında yıllık ücret talep edilmektedir.

İş zekâsı araçlarının tedariklerinde 3 seçenek öne çıkmaktadır [2]:

1. Ticari hazır araçlar,
2. Firma içerisinde geliştirilmiş araçlar,
3. Açık kaynak kodların belirli organizasyonu (pipeline) ile oluşturulmuş araçlar.

Bu seçenekler yukarıda belirtilen kriterler göz önüne alınarak mukayese edildiğinde Tablo-1'de verilen durum ortaya çıkmaktadır. Ticari ürünler yüksek seviyede hız, erişilebilirlik ve kaliteli teknik desteğe sahip olmasına rağmen lisans ve idame maliyetleri oldukça yüksektir. Öte yandan, ticari araçların esnekliği ve özelleştirilme işlevleri orta ve düşük seviyede olabilmektedir [5, 6, 7].

İş zekâsı çözümü	Esneklik	Hız	Erişilebilirlik	Teknik destek	Maliyet
Ticari araç	Düşük/orta	yüksek	yüksek	yüksek	çok yüksek
Şirket içinde geliştirilen araç	çok Yüksek	yüksek	yüksek	orta/yüksek	yüksek
Açık kaynak kod tabanlı araçlar	Orta/yüksek	yüksek	yüksek	orta	düşük

Firma içerisinde geliştirilen yazılımların özelleştirilebilirliği (ing: customization) oldukça iyi olmasına rağmen geliştirilme ve idame maliyetleri göreceli olarak yüksek olabilmektedir [8]. Bu araçların idame ettirilmesi, erişilebilirliği ve teknik desteğini sağlamak üzere vardiyalı ekip bulundurma

gerekebilmektedir. Bu şekilde geliştirilen ürünlerde de lisanslı veya açık kaynak kodlu yazılım kütüphaneleri kullanılması gerekmektedir. Zaman içerisinde bu kütüphanelerin teknoloji seviyesi düşük kalmakta olup güncellenmediği veya desteklenmediği takdirde yeni bilişim ürünleri ile uyumluluk göstermemektedir. Bu olasılık, firma ürünlerinin sürekli şekilde geliştirilmesini ve iyileştirilmesini gerekli kılmaktadır. Uzun vadede bu durum maliyetleri yükseltmektedir. Taşımacılık sektöründe de iş zekâsı araçları ve raporlamalar yönetim açısından oldukça önemlidir [9]. En güncel verilerle yapılacak analizler, geçmiş verilerden geleceğin tahminlendiği hızlı ve esnek sistemler üzerinde çalışan raporlar karar destek süreçlerinde yöneticilerin doğru kararlar almalarında önemli bir yer almaya başlamıştır. Ticari iş zekâsı ürünleri uyguladıkları kullanıcı bazında lisans, işlemci bazında lisans vb. fiyatlandırmalar ile oldukça maliyetli olabilmektedir [10]. Yurtiçi Kargo Servisleri AŞ özelinde ortaya çıkan maliyetler sebebiyle firma içerisinde yaklaşık 150 kullanıcıya bu hizmetler açılabilir. Halbuki bu raporların sahadaki yöneticiler ve şubelerle birlikte yaklaşık 1000 kullanıcıya açılması gerektiği değerlendirilmiştir. Bu miktarda kullanıcı için hesap edilen lisans maliyetleri sürdürülebilir nitelikte görülmemiştir. Bu sebeple firma içerisinde bir araç geliştirilmesi için proje başlatılmıştır. Tüm birimler tarafından kullanılan Misport isimli iş zekâsı ve raporlama aracı geliştirilmiştir. Ancak, bu aracın eski bir alt yapıyla geliştirilmiş olması, yeni teknolojilere destek vermemesi, güncel veriler ile raporlama imkanı olmaması ve idamesi için gerekli personel eksikliği sebepleriyle yeni bir raporlama aracının arayışına başlanmıştır. Yeni raporlama aracında olması beklenen özelliklerin başında yeni teknolojileri desteklemesi, güncel verilerle (gerçek zamanlı) raporlama imkanı sunması, rapor görselleri ile son kullanıcıya hitap etmesi, mobil araçlara uyumlu olması ve lisans maliyetinin olmamasıdır. Proje çıktısı olarak hedef kitlenin firma verilerini tek bir platform üzerinden izlemeleri böylece verimliliğin artması amaçlanmıştır.

## 2. Materyal ve Yöntem

Uçtan uca kullanılan Misport raporlama aracı ve diğer raporlama araçları tüm grup şirketleri bazında incelendi. Yapılan inceleme sonucunda veri büyüklüğü/kullanıcı sayısı fazlalığı açısından ilk aşamada yaklaşık 14 şirketten sadece Yurtiçi Kargo'ya ait raporların yeni raporlama aracına taşınmasına karar verildi. Geliştirme yapılacak platformu belirlemek için 5 açık kaynak iş zekâsı ürünü hız, esneklik, rapor görselleri, mobil uyumluluk ve duyarlı (ing: responsive) web tasarımı beklentisi açısından incelenerek Kibana üzerinde geliştirme yapılmasına karar verildi. Yurtiçi Kargo'nun karar destek süreçlerinde satış verisi önemli bir paya sahip olması nedeniyle satış verisinin Faz 1'de modellenmesi için çalışmalara başlandı. İş birimiyle birlikte satış raporları analiz edilerek hangi raporların taşınacağı, hangilerinin ekleneceği tespit edildi. Böylece oluşturulacak veri modeli için kullanılacak kaynak sistemler ve kaynak tablolar çıkarıldı. Uygulamanın yakın gerçek zamanlı bir sorgulama ve analitik platformu olmasından dolayı, verilerin zamanında ve tutarlı bir şekilde hedef sistemlere aktarılması önemli gerekliliklerden bir tanesidir. Kaynak sistemde üretilen verinin 1-2 dk içerisinde hedef sistemde sorgulamaya hazır hale getirilmesi hedeflenmiştir. Veri aktarımı için kullanılan modüller ve çalışmaların teknik mimari çizimi Şekil 1 de

verilmiştir.

## 2.1 Yapılan Çalışmaların Detayları

### 2.1.1 Verilerin veri kaynaklarından lisanslı bir ürün ile DWH ana veri tabanına alınması.

Lisanslı aktarım ürünü kaynak sistemlerden değişen verinin tespit edilip hedef sisteme aktarılması rolünü üstlenmektedir. Aynı zamanda mevcut süreçte lisanslı aktarım ürünü t-1 zamanda çalışan veri ambarı sistemimize veri sağlayan bir uygulamadır. Kullanılan lisanslı aktarım ürünü, kaynak sistemdeki veri değişikliklerini belirleyerek stage alana aktarmaktadır. Aktarılan veri hem veri ambarı hem de geliştirilen açık kaynak iş zekâsı ürününde kullanılmaktadır. Bu aktarım sırasında gecikmeler veya teknik problemlerin yaşanmaması için lisanslı aktarım ürününün kendine ait izleme (ing: monitoring) araçları üzerinden veri aktarımı sürekli izlenmektedir. İlerleyen zamanlarda lisanslı aktarım ürünü yerine sistem dahilinde kullanılan Kafka Cluster ile uyumlu çalışan Kafka Connector'ın kullanılması planlanmaktadır. Kafka Connector geliştirmesine devam edilmektedir.

### 2.1.2 Belirlenen verilerin SQL ve PLSQL yardımıyla denormalizasyon süreci içerisinde değerlendirilip aktarıma uygun hale getirilmesi.

Denormalizasyon sürecinde dahil olan işlemlerin tanımında eksiklikler yapılmaması ve kurgunun hatalı oluşmaması için belirlenen senaryo ve kurgular iş süreçlerinde kullanılan modeller baz alınarak çıkarılmıştır. Denormalize olacak verinin fazla olması durumunda verinin otomatik olarak bölünerek işleme tabi tutulması ile performans artışı sağlamıştır.

### 2.1.3 Denormalize olan verilerin Kafka Cluster'a aktarılması

Denormalize olan veri Java programlama dili ile geliştirilen Kafka Producer uygulaması ile Kafka Cluster'a gönderilmektedir. Oluşan verinin Kafka Cluster'a aktarımı için yazılmış olan Kafka Producer uygulamasının hatalı

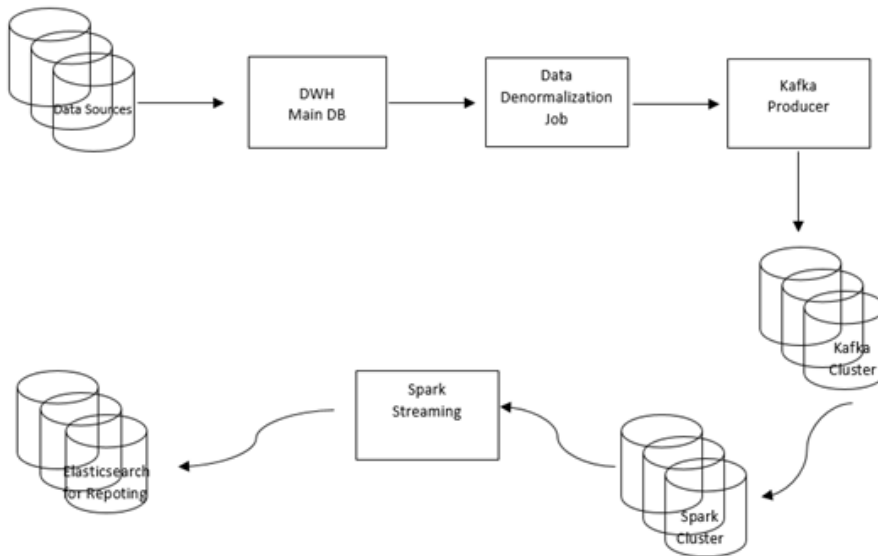
çalışması sonucunda eksik veya fazla veri oluşmaması için veriler zaman damgası ile işaretlenmiş ve en son okunan zaman bilgisi tutularak mükerrer okuma engellenmiştir.

### 2.1.4 Kafka Cluster a gelen verinin hesaplamalar için Spark Streaming e gönderilmesi.

Kafka Cluster da toplanan veri Java programlama dili ile geliştirilen Spark Streaming uygulaması ile Spark Cluster da işlenmektedir. Bir dakikalık sorgu aralıkları ile gelen veri üzerinde hesaplama yapan Spark Streaming işleminin zaman zaman donma (ing: hang) durumuna düşmesi, cevap verememesi gibi problemlerin yaşanmaması için Spark Streaming işleminin kullandığı bellek, merkezi işlem birimi (cpu) değerleri ve üzerindeki görevler izlenerek en uygun değerler ile çalışması sağlanmıştır.

### 2.1.5 Spark Streaming de üretilen hesaplanmış verinin sorgulamanın yapılacağı hedef veri tabanına aktarılması.

Verinin son hali Spark Streaming ve Spark - Elasticsearch uygulama protokol arayüzleri (ing: application protocol interface - API) ile Elasticsearch'e yazılmaktadır. Hedef olarak kullanılan veri tabanlarından bir tanesi olan Elasticsearch'de bir tekil alan kullanımı olmadığından, mükerrer verilerin gelmesi durumunda verinin çoklanması ve yanlış sonuçlar vermesi durumu oluşabilir. Bu adımdan önce verinin mükerrer okunması durumu oluşsa bile Elasticsearch üzerinde mükerrer oluşmaması için verinin yapısına göre üzerinde kullanılabilir tekil alanlar belirlenip aynı veriden tekrar oluşmaması sağlanacaktır. Veri modelinin oluşturulmasından sonra satış raporlarının görselleri ve kumanda tabloları Kibana arayüzünde hazırlandı. Veri testi için verilerin tüm kullanıcı grupları için veri ambarı ile karşılaştırılmasına, son kullanıcı testlerinin genel müdürlükte 36 pilot kullanıcıyla yapılmasına ve 1000+ kullanıcıya sistemin ne ölçüde cevap vereceğini görmek için kullanıcı yük testlerinin yapılmasına karar verildi.



Şekil 1. Teknik Mimari

### 3. Bulgular

36 pilot kullanıcıyla birlikte tüm kullanıcı grupları için yapılan veri testi sonucunda veri modeli üzerinde herhangi bir tutarsızlık ile karşılaşılması. Bunun en önemli sebebi veri modeli hazırlanırken genel müdürlük kullanıcılarının karar destek süreçlerinde hali hazırda kullandığı veri ambarı satış modelinin baz alınmasıdır. Veriler kaynak sistemden hedef sistem olan Elasticsearch veri tabanına 1 dk aralıklarla toplanarak (ing: aggregate) özetlenmiş olarak alınmaktadır. Verinin alınırken üzerinde tekil anahtar (ing: primary key) alanının olmaması sebebiyle veri aktarımında bir sorun yaşanırsa veri kirliliği sorunu ile karşılaşılabilir. Proje çıktısının kullanıma açılmasından itibaren %99 başarıyla çalışan aktarımda bu tip bir sorunla karşılaşılırsa kaynak ve hedef sistemi karşılaştıran kontrol sorguları hazırlanmış olup farklılık gösteren veriler tekrar aktarılmaktadır. Gelecek dönemde verinin tekil anahtar yapısıyla alınabileceği bir alt yapının kurularak aktarılması hedeflenmektedir. Kullanıcı yük testleri eşzamanlı 1000+ kullanıcının sisteme giriş yaptığı varsayımı ile tüm zaman filtreleri için yapıldı. Mevcut donanım alt yapısıyla anlık(bugün) sorgulamalarda sistemin 1-2 sn. içerisinde cevap verdiği ancak zaman filtresi genişletildiğinde (1 hafta, 1 ay, ...) veri kümesi de büyüdüğünden dolayı sistemin cevap verme süresinin uzadığı tespit edildi. Zaman filtresi 1 yıl olarak sorgulandığında ise mevcut donanım alt yapısıyla sistemin cevap verme süresinin 5 dk yı aştığı ve zaman aşımına (ing: timeout) düştüğü belirlendi. Bu sebeple faz 1'de 1 yıl zaman filtresi sadece üst yönetim kullanıcıları için tanımlandı. Hedef sistem üzerinde verilerin silinmemesi verinin büyümesine neden olmakta ve sonucunda büyük tarih aralıklarındaki sorgulamalarda performans sıkıntılarına sebep olmaktadır. Bu sebeple daha performanslı sorgulamaların yapılabilmesi için altyapının kurulması ilerleyen dönemlerde hedeflenmektedir. Açılması planlanan 1000+ kullanıcının kullanıcı grupları belirlenerek ERP ve Ldap senkronizasyonları tamamlandı. Uygulama seviyesinde güvenlik yönetimi kullanıcı ve rol bazında yapıldı. Kullanıcıların görmeye yetkili olduğu veriler hem rollerine hem de buldukları hiyerarşiye göre filtrelendi. Canlıya geçişte 1000+ kullanıcı kümesinin içerisinde ilk genel müdürlük kullanıcılarına 1 ay süreyle uygulama kullandırılarak sistemden ve kullanıcılardan alınan geri dönüşler iyileştirildi. Ardından bölge müdürlüğü ve birim kullanıcılarını içeren diğer kullanıcı gruplarına da açılarak canlıya geçiş sağlanmış oldu. Açık kaynak ürün geliştirmelerinde yaşanabilecek bir problem, lisanslı ürünlerde olduğu gibi kurumsal bir firmadan destek alınmıyor olmasıdır. Bu sebeple teknik geliştirme sürecinde yaşanan sıkıntılar mevcut iç kaynaklar ve internet üzerinde bulunan konu ile ilgili topluluklar vasıtasıyla giderilmiştir.

### 4. Tartışma ve Sonuç

Bu çalışmada açık kaynak kodlu teknolojiler ile mevcut iç kaynaklarla geliştirilen yakın gerçek zamanlı iş zekâsı uygulaması ve elde edilen sonuçlar incelenmiştir. Projenin yapılmasıyla uçtan uca açık kaynak iş zekâsı ürünü geliştirilmiş olup üst yönetimden birimlerimize kadar yaklaşık 1000+ kullanıcının karar destek süreçlerine destek sağlanmıştır. Mevcut kaynaklarla geliştirilen ürün şirkete hem geliştirme maliyeti hem de kullanıcı lisans maliyeti açısından ciddi avantaj sağlamıştır. Proje çalışmalarıyla iç kaynak personelin açık kaynak teknolojilerdeki yetkinlikleri artmış, büyük veri projeleri için gerekli alt yapı kurulmuştur. Proje sayesinde t-1 gün raporlamadan kurumun en büyük ihtiyaçlarından biri olan yakın gerçek zamanlı analitik veriye erişim sağlanmış, karar destek süreçlerinde ilgili birim

yöneticilerine destek olunmuştur. Uygulamanın canlıya alınması sonrasında kullanıcılardan olumlu geri dönüşler alınmış olup raporlama aracı giriş sıklığının proje öncesinde kullanılan raporlama aracına göre arttığı tespit edilmiştir. İlerleyen süreçte tahminleme modellerinin geliştirilmesi ile karar destek süreçlerine katkı sağlanması hedeflenmektedir.

### Kaynaklar

- [1] Ranjan, J. (2009). Business intelligence: Concepts, components, techniques and benefits. *Journal of theoretical and applied information technology*, 9(1), 60-70.
- [2] Gina, B., & Budree, A. (2020). A Review of Literature on Critical Factors that Drive the Selection of Business Intelligence Tools. In *2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)* (pp. 1-7). IEEE.
- [3] Sandu, D. I. (2008). Operational and real-time Business Intelligence. *Revista Informatica Economică*, 3(47), 33-36.
- [4] Horakova, M., & Skalska, H. (2013). Business intelligence and implementation in a small enterprise. *Journal of systems integration*, 4(2), 50-61.
- [5] Lapa, J., Bernardino, J., & Figueiredo, A. (2014, May). A comparative analysis of open source business intelligence platforms. In *Proceedings of the International Conference on Information Systems and Design of Communication* (pp. 86-92).
- [6] Vargas, V., Syed, A., Mohammad, A., & Halgamuge, M. N. (2016). Pentaho and JasperSoft: a comparative study of business intelligence open source tools processing big data to evaluate performances. *International Journal of Advanced Computer Science and Applications*, 7(10), 20-29.
- [7] Thomsen, C., & Pedersen, T. B. (2005, August). A survey of open source tools for business intelligence. In *International Conference on Data Warehousing and Knowledge Discovery* (pp. 74-84). Springer, Berlin, Heidelberg.
- [8] Habryn, F., Kunze von Bischoffshause, J., & Satzger, G. (2012). A business intelligence solution for assessing customer interaction, cross-selling, and customization in a customer intimacy context.
- [9] Jin, D. H., & Kim, H. J. (2018). Integrated understanding of big data, big data analysis, and business intelligence: a case study of logistics. *Sustainability*, 10(10), 3778.
- [10] Nofal, M. I., & Yusof, Z. M. (2013). Integration of business intelligence and enterprise resource planning within organizations. *Procedia technology*, 11, 658-665.