

Araştırma Makalesi

İstatistik Yöntemleri ve İstatistik Mühendisliğinin Dil ve Edebiyat Araştırmalarında ve Mühendislik Alanında Kullanımına İlişkin Önerilere Bir Bakış

Adnan Mazmanoğlu^{1,*} 

¹ Endüstri Mühendisliği, Mühendislik Fakültesi, Toros Üniversitesi, İstanbul, Türkiye

¹ adnan.mazmanoglu@toros.edu.tr

Geliş: 08.08.2023

Kabul: 25.12.2023

DOI: 10.55581/ejeas.1339553

Öz: İstatistikçiler, sundukları yöntem (metodoloji) desteği dolayısıyla diğer bilim dallarına araştırma, planlama, veri toplama ve analiz aşamalarında yardımcı olur. Bu bakımdan bazen değişik disiplinlere ait uygulamalı çalışmalar yapar veya araştırma gruplarında yer alırlar. Çağımız bilimine yön veren ülkelerde durum budur. Bizde de artık aynı yola gidişin işaretleri belirmektedir. Böylece, bir zamanlar çuvallar dolusu fiş yazmak zorunda kalan edebiyatçılarımız, şimdi bilgisayar programları yardımıyla metinler üzerinde indeks çıkarma, kelime sayma, alfabetik sıraya koyma gibi işleri anında yapabileceklerdir. Araştırmacı için bir kâbus olan ve yıllar alan bu mekanik işlemler artık birkaç dakika içinde çözülmekte, insan zekâsı asli görevi olan düşünme faaliyeti için serbest kalmaktadır. Öte yandan, saniyede bir milyarları bulan işlem hızıyla bilgisayarlar, kâğıt kalemle asla yapılamayacak veya yıllar alacak hesapları hatasız şekilde sonuçlandırarak, dağınık ve karmaşık bilgi yığınlarından (bilhassa büyük veri~ big data) mantıklı teorilere yol açacak ana fikirlerin süzülmesini sağlamaktadır. 19.yüzyıldan bu yana istatistik analiz yöntemleriyle dil ve edebiyat üzerinde çalışmalara ve istatistik mühendisliği sayesinde mühendisliğin her alanında veri olan her yerde yer verilerek hangi amaçla yapıldığı ve sağlanan yararlar üzerinde örneklerle durulmuştur. Bilhassa istatistiğin ve istatistik mühendisliğinin diğer kullanım olanakları üzerine öneriler verilmiştir.

Anahtar kelimeler: Entropi, İstatistik Mühendisliği, İstatistik Yöntemi, Metin Karakteristiği

A Survey of the Proposals for the Use of Statistical Methods and Statistical Engineering in Language and Literature Studies and Engineering

Abstract: Statisticians assist other disciplines in research, planning, data collection and analysis stages due to the methodological support they offer. In this respect, they sometimes carry out applied studies from different disciplines or take part in research groups. This is the situation in the countries that direct the science of our age. There are signs of going the same way for us now. Thus, our literati, who once had to write sacks full of receipts, will now be able to instantly perform tasks such as indexing the texts, counting words, and putting them in alphabetical order with the help of computer programs. These mechanical processes, which are a nightmare for the researcher and take years, are now solved in a few minutes, and the human mind is freed for its main task, the thinking activity. On the other hand, with a processing speed of up to one billion per second, computers allow to filter out the main ideas that will lead to logical theories from messy and complex piles of information (especially big data) by concluding calculations that could never be done with pen and paper or take years. Since the 19th century, studies on language and literature with statistical analysis methods and statistical engineering have been included in every field of engineering, where data is available, for what purpose it is done and the benefits it provides are emphasized with examples. In particular, recommendations are given on other uses of statistics and statistical engineering.

Keywords: Entropy, Statistical Engineering, Statistic Methods, Word characteristic

*Sorumlu yazar

E-mail Adress: adnan. mazmanoglu@toros.edu.tr (A. Mazmanoğlu)

1. Giriş

Önce sosyal bilim konularından örneğin Türk Dili ve Edebiyatı'nda zaman zaman istatistik dergilerde bu yönde yapılmış yayınlara rastlanmaktaydı. Türk Dili ve Edebiyatı'nda bu çalışmaların yapılmaması için hiçbir sebep yoktu. İlk olarak, 1973 yılında Türk Roman'ındaki Batı kaynaklı kelimeler üzerinde bir çalışma başlatıldı. Bu arada, çeşitli metinlerde kelime uzunluklarını ele alan bir araştırma 1979 III. Milletler Arası Türkoloji Kongresi'nde bir bildiri halinde sunuldu. Bunu 1982 ve 1985 kongreleri takip etti. 1982 bildirisinde, romanımızdaki batı kaynaklı kelimelerin zaman içinde gittikçe yoğunluk kazandığı rakamlarla ispat edilmekte, dil ve kültürümüzde batılılaşmanın yarattığı kutuplaşmalar ilmi metotlarla ortaya konmaktaydı. Bu bildirinin sunulmasından sonra dinleyicilerden gelen sorular ve yapılan tartışmalar üzerine bir açıklama yapılması gerekli oldu. "İstatistik ve Edebiyat" başlıklı bu uzunca makalenin ancak bir kısmı 1983 yılında Türk Edebiyatı dergisinde yayımlandı. Daha sonra, 1984 Yazarlar Birliği Yıllığı'na da alınan bu makalenin tamamı ilk olarak bu kitapta yer almaktadır. 1985 Türkoloji Kongresi'nde sunulan "Dil ve Edebiyatta Ölçülebilir Özellikler" başlıklı bildiri, edebiyatçılarımız için niceleyici çalışmalarda yol gösterici olabilir.

2. Dil ve Edebiyatıta Ölçülebilirlik

2.1. Çağdaş Anlamda İstatistik

Çağdaş anlamda istatistik, doğal olaylara dayanan ve gözlemlenilen verilerin bilimsel yöntemlerle incelenmesi ve doğru sonuç çıkarılmasıdır. Genel olarak; gözlemlenilen verileri düzenleme, analiz etme ve bunlardan sonuç çıkarma sanatı ve bilimi olarak istatistik ortaya çıkar. Sayısal bilgiden anlam çıkarma, belirsizliğin ele alınması, örneklemenin önemi, tahmin yapma ve karar verme bağlantılarını geliştirmek için özellikle rastgele deney ve gözlemlere dayalı bilimsel ve işlevsel bağlantıları kullanmak çok önemlidir. Zira veri gruplarının düzenlenmesi, tanımlanması, sunulması ve grafiklerle, listelerle, tablolarla ifadesi bu becerilerin kazanılması çağdaş istatistikçinin yeni istatistik mesleklerinden "veri bilimcisi" donanımını elde etmek için olmazsa olmaz becerilerdir [1].

İstatistik yöntemleri edebiyatta diğer yorum şekillerinin yanında onlarla aynı değerde ve aynı düzeyde değil, metinlerin şekil özelliklerini betimleyen yardımcı unsurlardır. Hatta istatistik yöntemlerinin teolojide (manevi bilimlerde) kullanılması, bugün bile hala büyük bir şüphe ile karşılanmaktadır. Üniversitelerin araştırma ve bilgi değerlendirme merkezlerinde istatistik sonuçlarının gösterdiği gibi sosyologlar, psikologlar ve dilbilimciler, istatistiği uzun zamandan beri kullanmaktadırlar [2].

İstatistiksel edebiyat analizlerinin asıl problemi, onun yayıldığı alandır. İstatistik açısından elde edilen sonuçlar ile bir metnin anlamı içeriği ve eserin istatistik değeri arasında benzerlikler var mıdır? Dış görünüş özelliklerini anlam biliminde de etkili olması, onların anlam bilim ölçülere uyması mümkündür. Yeterli sayıda eser açısından bunu belgeleyebilen araştırmalar ise şimdiye kadar var olmamıştır. İstatistik

verilerin değerlendirmede, bu veriler arasında bir ilişki, bir bağlantı kurmak daha da zordur. İstatistik yönünden geleneksel ve öznel (sübjektif) olarak elde edilen bilgilerden daha doğru ve estetik problemler için büyük önemi olan tek ölçü olarak bir eserin tekdüze (homojen) oluşu anımsanabilir [2].

Jonathan Swift, 1725 yılında "Julver'in Seyahatleri"inde Laputan dilinin genel kelime kaynağını içeren ve bunun yardımıyla bu konudaki tüm hesaplamaları verebilen bir makineyi tasvir eder, "ayrıca nesnelere, fiillerin, isimlerin ve diğer cümle öğelerinin sayısı arasında kitaplardaki ilişkiler de verilmektedir" [2]. Jonathan Swift, burada kendi çağının bilimini uğraşı terapisi hatta dil bilimcileri de hece sayıcıları olarak hicveder.

Dildeki gelişmelerin betimlenmesinde kullanılan sayısal analiz yöntemleri, ciddi şekilde 19. Yüzyılın ortalarında denenmiştir. Fakat istatistiksel dil analizleri, ancak bilgisayarlardaki bellek kapasitesinin daha büyük hacimli veri değerlendirme etkilerinin hazırlanmasından yakın zamanda kolay ve hızlı bir şekilde önem kazanmıştır.

İçinde bulunduğumuz 20. yüzyılın ancak altmışlı – yetmişli yıllarından bu yana bir eserin veya bir edebi türün şekil özellikleri ve kullanılan kelimeler hakkındaki varsayımlar ~ iddialar ~ hipotezler, istatistik konulu metinler tarafından doğru olarak test edilip denetlendikten sonra ret veya kabul edilebilmeye başlanmıştır. Manon MAREN – Grisebach'ın yazdığı Edebiyat Bilimi'nin Yöntemleri kitabında (Türkçe 'ye çeviren Dr. Arif Ünal) öğrendiğimize göre 1852 yılında E. Förstermann'ın Almanca 'da, Latince 'de ve Yunanca 'da Sayısal Dil İlişkileri (münasebetleri)" adlı eseri yayımlanmıştır. Bu biçimdeki çalışmalar, 1890'lı yıllarda P. Valerio ve F. W. Käding tarafından devam ettirilmiştir.

Yeni çalışmaların öncüleri olarak E. ve K. Zwirner'in değişik dillerdeki ses ilişkileri hakkındaki çalışmaları ve G. K. Zipf'in, "dil'in Fizyolojisi" adlı eseri sayılabilir. Zipf, ilk önce bir metnin kelime hazinesindeki bir kelime ile çok kullanılan kelimeler sözlüğündeki bu kelimenin önceliğinin (r) sık sık kullanılışı arasındaki ilişkiye (yani istatistiksel anlamda korelasyona) işaret etmiştir. Ayrıca tekrarlanma sayısının ölçüsünü bulan frekans dağılımı ile burada geçen sözlükte en sık geçen kelime r1, en sık geçen kelime r2 vb. ile gösterilerek cümle uzunluğu hakkında en geniş bilgiyi ilk önce G. U. Yule ve C. B. Williams vermişlerdir [3].

2.2. Metin Karakteristikleri

Edebi eserlerin metin karakteristiğinin saptanması, ilk önce Achen'li fizikçi Wilhelm Fucks tarafından yapılmıştır [4]. Fucks, harfler, heceler, cümleler, bunların özellikleri ve birbirleriyle ilişkisi gibi şeylerden oluşan bir metnin unsurlarını betimlemektir. İstatistiksel bir metnin analizinin amacı, bir metnin dış görünüşü hakkındaki fikirlere. Gerekli bir tanımlamanın (description) amacı olarak bu işlemler yalnız, ağırlık noktası dış görünüş yönünden bulunan metinlerde söz konusu olmaktadır. Bu tür eserlerin genel edebiyata katkıları ölçüsünde istatistiksel inceleme yöntemi önem kazanır. 20. Yüzyılın başından beri bu alanda bir artış olduğu; (1) bilgisayar sistemlerinin bellek ve işlem hızının çok yüksek

düzye de olması, (2) bu gelişmeye paralel olarak istatistik bilimindeki analiz yöntemlerinin her veriye uygun olarak geliştirilmesi ve yazılımlarının hazırlanmasından dolayı tartışılmaz. Bu konuda Manon MAREN – Grisebach’ın bize ulaştırdığı çok örnek olmakla birlikte bazılarını sizlerle paylaşalım [2]. Christian Morgenstern [6]’in metinleri, bilhassa August Stramm’ın [7] ve Dadaistler’in çalışmaları [8] ve İngiliz Edebiyatı çerçevesinde E. E. Gumming [9] ve Getrude Stein’in [10] çalışmalarından söz edilebilir. Helmut Heissenbuttel [11] ve Fransız Cumhuriyetçi yazarların çalışmaları, günümüz edebiyatından örnekler sunar. Edebiyatın yaygın yorum yöntemleri çerçevesindeki sık sık bir analizle anlaşılacak metinlerle karşılaşırız. Örneğin Wilhelm Fucks [4], metnin hecelerinden anlaşılan özel biçim özelliklerini araştırır: Her kelimedeki orta hece sayısı ve bundan hesaplanan ve “Metin Entropisi ~ incelenen metnin unsurlarının düzen derecesini verir”. Diğer metin karakteristikleri (hecelerde), orta cümle uzunlukları ve bir metindeki aynı hece sayısının kelime zincirinin orta uzunluğudur. İstatistiksel incelemeler, ayrıca metrik analizlerini de vermektedir.

2.3. Entropi

Bir sistemdeki belirsizliğin ölçüsüne “entropi” adı verilir. S bir kaynak olsun. Bu kaynağın $\{m_1, m_2, \dots, m_n\}$ olmak üzere n mesaj (ileti) üretilebildiğini varsayalım. Tüm mesajlar birbirinden bağımsız olarak üretilmektedir ve m_i mesajlarının üretilme olasılıkları p_i ’dir. $P = \{p_1, p_2, \dots, p_n\}$ olasılık dağılımına sahip mesajları ~ metinleri üreten S kaynağının entropisi $H(S)$ şu şekildedir [12]:

$$H(S) = -\sum_{i=1}^n p_i \log_2(p_i) \quad (1)$$

Örneğin sekiz kişinin bir soruya verdiği yanıtlar S kümesindeki gibi olsun.

$S = \{\text{evet, evet, hayır, hayır, hayır, hayır, hayır, hayır}\}$

Olasılıklar, iki adet “evet” değeri için,

$$p_1 = 2/8 = 0,25 \quad (2)$$

diğer altı adet “hayır” değeri için,

$$p_2 = 6/8 = 0,75 \quad (3)$$

hesaplanır. O halde $P = (p_1, p_2) = (0,25, 0,75)$ yazılır ve bu değerlere bağlı olarak S için toplam entropi aşağıdaki şekilde elde edilir.

$$H(S) = -[p_1 \log_2(p_1) + p_2 \log_2(p_2)] \quad (4)$$

$$= -[0,25 \log_2(0,25) + 0,75 \log_2(0,75)]$$

$$= 0,81128 \approx \%81$$

Yani S kümesine hangi yanıt geleceğinin belirsizlik ölçüsü %81 bulunmuştur denir. Bir bilgi sisteminin toplam enformasyon (bilgi ~ malumat) potansiyeline, Shannon, o sistemin “entropisi” adını vermiştir. Hatta metin çözümlemede daha çok termodinamikten alınan bu terimi ise kendisine büyük matematikçi ve bilişim kuramcısı John Von Neumann ısrarla önermiş ve “Entropinin ne olduğunu kimseler anlamaz, böylece sen de enformasyon kuramını anlatırken, hem ne

kadar derinmiş diye düşünürler, hem de söyleyecek pek bir şey bulamaz” demiştir. Bazı uzmanlar, Claude Shannon’un enformasyona etkisini, alfabenin mucidinin edebiyata etkisine benzetecek kadar önemli bulmaktadırlar. Hatta günümüzde kompakt disklerden süper bilgisayarlara, internetten uzay araçlarına kadar, sayısal bilgiyi saklayan, işleyen ve ileten tüm sistemlerin temelleri Shannon’un enformasyon kuramına dayanmaktadır. Şimdi yine metin entropisine uygun yapılan araştırmalardan söz etmeye çalışalım [13].

Hecelerdeki orta cümle uzunlukları için birkaç örnek verelim: Rilke’nin “Cornet” i 12.5, Goethe’nin İtalya Seyahati 38.8, Marx’ın “Kapital”i 65.4 ve Bild Gazetesinin 1968 yılı Nisan ayındaki birkaç makale 20.5.

Farklı türlerin ve farklı yazarların metinlerinin incelenmesi, metin benzetmelerinin saptanmasının sonuçlarındaki rastgele farklılıklardan kurtulmak için yaklaşık 300-400 kelime uzunluğunda bir metnin yeterli olduğu sonucunu ortaya çıkarmıştır. Bu metnin farklı kısımlarının değerler dağılımı, dış şekil birliği için bir ölçü verir. Sayımların kesin sonucu, bir eserin bölümlerindeki özellikler açısından benzerliklerin ve farklılıkların saptanmasıdır. Bunun da yazarın diğer eserlerinin ve diğer yazarların eserlerinin özellikleri karşılaştırılarak yapılan çalışmalarda değer “sapma (Standard deviation)” larıyla bağlantısı vardır. Örneğin bir nesir metninde ortalama hece sayısı için değerlerin sapmasında her kelime için 0.01 – 0.02 normal karşılanmıştır. Bölümdeki sapmalarda dağılımdan bağımsız olarak genellikle tüm metnin ortalama değerinden kelime başına 0.2 hecelik bir sapma ortaya çıkar. Fucks, istatistiksel çalışmaların temelinde doğal dillerdeki kelime oluşturma teorisi ile matematiksel bir dil biçim teorisi geliştirmiştir.

2.4. İstatistiğin Diğer Kullanım Olanakları

İstatistiksel incelemeler, bir edebi türdeki metinlerin dış görünüşündeki ince farklılıkları göstermeye, şimdiye kadarki analizlerden çıkan yapı özelliklerini ispat etmeye, bir çağın (devrin) şekil özelliklerine uygun olarak bir eserin durumunu aydınlatmaya ve şekildeki gelişmelerin çizgisini belirlemeye ve tercüme hakkında bir karar varırken bazı kriterler saptamaya uygun görünmektedir.

Örneğin, verilerin kümeleneşi, analizi, nesneleri şiir, nesir vb. metinlerde birbirine olan benzerliklerine göre gruplarına ayırmak demektir. Elde bulunan veriler incelenerek birbirine benzeyenler bir kümeye, benzemeyenler ise başka bir kümeye toplanır. Verilerin kümeleme analizine göre modellenmesinde matematik, istatistik, bilgisayarlı makine öğrenimi ve yapay zekâ gibi birçok alandan yararlanılır. Bilhassa bu yöntemle ve benzer istatistik analiz yöntemlerle dil ve edebiyatta, Türkçe nesirlerde ve aruz vezninde kullanılan kelime uzunluklarından Türk romanındaki Batı kaynaklı kelimelere ve buradan hareketle yazarlarımızın sınıflandırma (kümeleme) denemesine kadar çalışmalar yapılabilir. Bilgisayar kullanımının yaygınlaşması hem hız hem de bilgi saklama kapasitelerinin artması, bilgisayar kullanımı ile bilgiye ulaşma ve onu işleme aşamasında bilim adamlarının önüne araştırma yapmak için sonsuz hedefler açmaktadır.

Oysa bilgisayarlar bilhassa metinler üzerinde indeks (fihrist) çıkarma, kelime sayma, düzeltme, sıralama, saklama, sorgulama vb. gibi işlemleri çok kısa bir (saniyeler

mertebesinde) zaman içinde çözümleyebilmektedirler. Bu açıklamaya ek olarak bilhassa istatistik 20 yüzyıldan itibaren geçirdiği yüksek gelişim, ortaya koyduğu veriye, konuya vb. uygun ölçme ve değerlendirme teknikleriyle edebiyat araştırmalarına yeni boyutlar katarak, onları zenginleştiriyor [14]. Biz istatistikçiler olarak bu çalışmada sunacağımız metodolojik destekle yani veri toplama ve analiz aşamasında dil ve edebiyata yardımcı olmaya çalışacağız. Çağımız bilimine yön veren söz sahibi ülkelerde durum böyledir. İstatistiksel edebiyat analizlerinin asıl problemi, onun yayıldığı alandır. Son yıllarda bir istatistik sistemi olarak "istatistik mühendisliği" kavramının ortaya çıkması, istatistiksel düşünme yoluyla bilim ve istatistik teorisi kullanılarak oluşturulan bu yeni istatistik disiplini sayesinde internetteki bilgi ve veri yoğunluğunun ve bunları depolayan makinelerin de eklendiği veri okyanuslarının doğru, analiz yapabileme gücü ancak aşılabilecektir. Büyük Veri (Big data) nın çok büyük miktarda bilgiyi öğretmek, onu anlak olarak analiz etmek ve kimi zaman ondan enteresan sonuçlar çıkarmak için yeni bulunmuş, yeni ortaya çıkan bir yaklaşım olduğunu söyleyebiliriz. Bu da istatistik analiz tekniklerinin ciddi olarak eğitiminin alınmasını gerekli kılmıştır.

3. Dil ve Edebiyatta İstatistik

İlk bakışta sanat ve edebiyatla istatistik ve bilgisayar arasında hiçbir ilgi yokmuş gibi gelebilir. Bazılarına göre sanat sadece hissedilebilir. Çünkü zevkler ve renkler tartışılmaz. Matematik ve buna dayalı, müspet ilimler ise aklın eseri olup objektiftirler. Kesin ispatlara dayanırlar. Sanatla ilmin malzemesi, yöntem ve ürünü tamamen birbirinden farklıdır. Dolayısıyla herkes kendi yoluna gitmelidir. Şair kelimeleriyle uğraşsın, mısralarını yontsun, cılasın. Matematikçi de formüller türetsin, denklemler kursun.

Çoğunluğun desteklediği bu düşünce tarzı ilk bakışta çekici görünmekte ise de esaslı bir temelden yoksundur. Duygu ve düşünceler arasında duvar örülemez. Güzellik içten içe bir ölçüye dayandığı gibi, dikkatle incelenirse matematik ve hendese (mühendislik) ruhunun da vardığı nokta estetik bir değer taşır. Aslında günümüzde her alanda kaydedilen baş döndürücü gelişmeler, birçok ilim ve sanat adamını dar sahalara içine hapsedmiştir. Daha XVIII. ve hatta XIX. asırda iyi eğitilmiş bir kişi, zamanın tüm bilimleri bilme imkânına sahipti. Bin yedi yüzlü yıllarda bir bilim adamı aynı zamanda bir filozof veya sanatçıydı. Yalnız bilgi değil, irfan ve hikmet sahibiydi. Astronomi, matematik, fizik, tıp, dilbilim, din, sanat, felsefe, müzik ve mimari hakkında zamanının bütün bilgilerine vakıf olabilirdi. Bu durum, Doğu ve Batı bilginlerinin ortak vasfıydı. Zamanımızın sanatçı ve bilim adamı neden böyle değil? Sorunun cevabı gayet açıktır: Hafıza, zekâ ve zaman bakımından sınırlı olan insan, değil bütün bilim ve sanat dallarında, tek bir dalın ince ayrıntısında dahi mevcut bilgilerin tamamını öğrenmekten aciz hale gelmiştir. Çeşitli dallardaki bilimsel gelişmeleri ele alan araştırmacılar, bunların zaman içinde üslû bir fonksiyon şeklinde arttığını ortaya koymuşlardır. Bilimdeki yenilikler her on yılda bir katlanmaktadır. Bu demektir ki, 1940 yılında kendi uzmanlık alanındaki bütün bilgilere sahip olan bir bilgin, 1950'deki bilgilerin ancak yarısına, 1960 dakilerin dörtte birine sahiptir. 1990 dakilerle kıyaslandığında, onun bildikleri otuz ikide bire düşmüş demektir. Bundan ötürüdür ki bir

yandan sürekli olarak yeni ilim dalları ortaya çıkmakta ve alt dallar oluşmaktadır. Gün geçmiyor ki insanlığı hayretler içinde bırakan yeni bir buluştan söz edilmesin. Henüz teknolojiye aktarılmayan bilgileri ise kütüphane rafları bile artık çekememektedir. 1970 li yıllarda İngiltere Patent Dairesi kütüphanesinde yer alan yazı ve dergilerde 320000 konu başlığının bulunduğu düşünülürse, birbirinden habersiz olarak ilmin derin dehlizlerinde yol alan milyonlarca araştırmacının, yaptıkları buluşlarla ortaya koydukları mevcut ilim birikimi karşısında her feridin cahillik derecesini gün geçtikçe derinleştirdikleri ortaya çıkar. Gelmiş geçmiş bütün ilim adamlarının %90 ı halen hayatta olmakla beraber, bu kişiler birbirinin alanından uzak ve habersizdirler [14].

İstatistik ise yüzyılımızın başından itibaren yaptığı hamlelerle, her bilim adamının yararlanacağı yöntemleri ortaya koymuş, ölçme ve değerlendirme teknikleri sunmuştur. Bununla beraber bu yöntemlerin yaygınlık kazanması da ancak bilgisayarlar sayesinde olmuştur. Dil ve edebiyatın, İstatistik metodları kullanması ve bilgisayarlardan faydalanması da artık kaçınılmazdır. Çağın gidişi bu yöndedir. Günümüzde dil ve üslubu ele alan Stylometry adlı bir İstatistik dalı gelişmiş, dilbilime uygulanan bilgisayar yöntemlerini inceleyen Computational Linguistics gibi dergiler yayımlanmaya başlamıştır.

Zaman zaman istatistik dergilerde bu yönde yapılmış yayınlara rastlanmaktaydı. Türk Dili ve Edebiyatı'nda bu çalışmaların yapılmaması için hiçbir sebep yoktu. İlk olarak, 1973 yılında Türk Roman'ındaki Batı kaynaklı kelimeler üzerinde bir çalışma başlatıldı. Bu arada, çeşitli metinlerde kelime uzunluklarını ele alan bir araştırma 1979 III. Milletler Arası Türkoloji Kongresi'nde bir bildiri halinde sunuldu. Bunu 1982 ve 1985 kongreleri takip etti. 1982 bildirisinde, romanımızdaki batı kaynaklı kelimelerin zaman içinde gittikçe yoğunluk kazandığı rakamlarla ispat edilmekte, dil ve kültürümüzde batılılaşmanın yarattığı kutuplaşmalar ilmi metodlarla ortaya konmaktaydı. Bu bildirinun sunulmasından sonra dinleyicilerden gelen sorular ve yapılan tartışmalar üzerine bir açıklama yapılması gerekli oldu. "İstatistik ve Edebiyat" başlıklı bu uzunca makalenin ancak bir kısmı 1983 yılında Türk Edebiyatı dergisinde yayınlandı. Daha sonra, 1984 Yazarlar Birliği Yıllığı'na da alınan bu makalenin tamamı ilk olarak bu kitapta yer almaktadır. 1985 Türkoloji Kongresi'nde sunulan "Dil ve Edebiyatta Ölçülebilir Özellikler" başlıklı bildiri, edebiyatçılarımız için niceleyici çalışmalarda yol gösterici olabilir.

4. İstanbul Edebiyat Fakültesinden Bir Makale: İstatistik Metod Eleştirisi Kuramı Teorisi Hakkında Kısaca Özet Bilgi

Bütün bilim ve sanat dallarında etkilerini gördüğümüz istatistik metod kuramı [15], edebî metnin ve edebiyatla ilgili her bilginin incelenmesinde zaman zaman karşılaşılan bir dizi teklif ihtiva eder. Özellikle "verilerin" çok olduğu durumlarda verilerin birbiri ile ilişkilerinin incelendiği araştırmalar, istatistik metodun tekniklerini kullanmayı gerektirmektedir.

İstatistik, matematik bilimi ile yakından alakalı bir bilim dalıdır. Verileri, matematiksel değerleri açısından inceler. Aritmetik ortalama başta olmak üzere sayısal veri grubunun birbirlerine göre durumlarını çok çeşitli açılardan değerlendirir. Ortalamalarını alır, genel ve özel sayısal

değerlerini hesaplar, grafikler çizer, yüzdelerini ve varyanslarını (standart sapmalarını) bulur ve bunlara benzer daha pek çok çalışma ile uğraşır.

İstatistik, bir anlamda deney ve gözlemlerin matematiğidir, dolayısıyla sayma ve ölçme ile elde edilmiş verilerle uğraşan bilimsel bir yöntemdir. Saymalar ve ölçmeler sonunda araştırmacı karar verme aşamasına gelir.

İstatistik, günümüzün çok unsurlu hayatı içinde hemen her parçanın birbiri ile ilişkisini sayısal değerlerle verebildiği için önemlidir. Yıllarca fen bilimlerinin ortak bir metodu olan istatistik, artık edebiyatta da kullanılmaktadır. Edebî metinlerin dil ve üslup çalışmaları, istatistik olmadan yapılamaz. Sanatkarların kavram dünyaları ancak istatistik verilerle ölçülebilir. Bir şiirdeki asonans ve aliterasyon çalışması edebiyata İstatistik metodun en yaygın uygulamalarından biridir.

Kafiye, vezin, nazım şekli, nazım birimi gibi şiirin teknik özellikleri, istatistik ile sağlam bir biçimde ölçülebilir. Bibliyografik künye bilgileri, istatistiğin sıralama, gruplama gibi teknikleri sayesinde düzenlenebilir. Veriler içinde sayı, yüzde ve dağılım gibi değerler başta olmak üzere istatistik metodun edebî metin incelemesinde önemli bir yeri vardır.

İstatistik, günümüzün çok unsurlu hayatı içinde hemen her parçanın birbiri ile ilişkisini sayısal değerlerle verebildiği için önemlidir. Yıllarca fen bilimlerinin ortak bir metodu olan istatistik, artık edebiyatta da kullanılmaktadır. Edebî metinlerin dil ve üslup çalışmaları, istatistik olmadan yapılamaz. Sanatkarların kavram dünyaları ancak istatistik verilerle ölçülebilir.

Bir şiirdeki asonans ve aliterasyon (mısra~ dize içinde ve komşu mısralar arasında bir ahenkle tekrarlanan ünlüler asonans, ünsüzler aliterasyon oluşturur) çalışması, edebiyatta İstatistik metodun en yaygın uygulamalarından biridir.

Kafiye, vezin, nazım şekli, nazım birimi gibi şiirin teknik özellikleri, istatistik ile sağlam bir biçimde ölçülebilir. Bibliyografik künye bilgileri, istatistiğin sıralama, gruplama gibi teknikleri sayesinde düzenlenebilir. Veriler içinde sayı, yüzde ve dağılım gibi değerler başta olmak üzere istatistik metodun edebî metin incelemesinde önemli bir yeri vardır.

İstatistiki incelemeler, bir edebî türdeki metinlerin dış görünüşündeki ince farklılıkları göstermeye, şimdiye kadarki analizlerden çıkan yapı özelliklerini ispat etmeye, bir devrin şekil özelliklerine uygun olarak bir eserin durumunu aydınlatmaya ve şekildeki gelişmelerin çizgisini belirlemeye ve tercüme hakkında bir karara varılırken bazı kriterler tespit etmeye uygun görünmektedir. [6; gerçeğine sadık kalınarak alıntı yapılmıştır].

Bu çalışma İstatistik metotlar ve bilgisayar kullanımıyla dil ve edebiyat alanında neler yapılabileceğine birkaç örnek vermiş sayılabilirse amacına ulaşmış olacaktır.

5. İstatistik Mühendisliği

İstatistiğin son zamanlarda kullanılan yöntemlerden biri olan "veri madenciliği" ne bir göz atalım. Kurumlarda biriken veri içerisinden kurum için yararlı olanlarını bulup ortaya çıkarma işine veri madenciliği adı veriliyor. Bu eylem bizi "İstatistik Mühendisliğine" götürmektedir. Veri madenciliğinin

kullanılmadığı alan yok gibidir. Pazarlama, bankacılık, sigortacılık ve elektronik ticaret vb. değişik türde kullanılabilir. Veri madenciliğinin veri kümeleme tekniği dil ve edebiyatta istatistiksel analiz yöntemlerinin kullanılmasına ışık tutacağı için bir örnekle bir giriş yapalım. Rastgele bir makine kullanımının öğrenimi açısından, her bir küme gizli bir örüntüyü (pattern) temsil eder ve uygulanan öğrenme ise bir denetimsiz öğrenmedir. Bir açıdan bakıldığında ise kümelemeyi "gizli örüntülerin" ortaya çıkarılması için uygulanan bir denetimsiz "öğrenme yaklaşımı" olarak tanımlanabilir [16].

Kümelemedeki öğrenmenin denetimsiz olmasının nedeni önceden belirlenmiş sınıfların olmayışıdır. Önceden sınıflar belirli olsaydı, zaten bu kümeleme değil, bir sınıflandırma "modeli" adını alacaktı. Önceden sınıflar belirli iken, yani kadın ve erkek, evli veya bekâr diye iki ayrı sınıf varken yapılan (algoritmik) öğrenmeye denetimli öğrenme; herhangi bir sınıf adı verilmeden yapılan öğrenmeye ise denetimsiz öğrenme denilir. Örneğin, veri tabanındaki kayıtlarda her kaydı yanında kadın veya erkek, evli veya bekâr bilgisi yazılıyor olsun, bu durumda veri tabanı üzerinde yapılan herhangi bir (kadın veya erkek olduğuna ilişkin) kural çıkarma işlemi denetimli öğrenmedir. Ancak, yapılan kural çıkarma işlemi denetimsiz öğrenmedir. Bu işlem aynı zamanda veri tabanını (iki) kümeye ayırma, yani kümeleme işlemidir. Burada, kadın/erkek gibi bir etiket ya da sınıf olmayacağı için kümeleme kayıtlar arasındaki benzerlik veya uzaklık (mesafe) ölçütüne göre yapılır.

İki verinin (metnin, şiirin vb.) benzerliğinden kasıt ise aralarındaki uzaklığın ölçülmesi ve değerlendirilmesidir [17]. Bu değerlendirme veri tabanındaki diğer verilere oranla iki verinin ne kadar yakın ya da benzer oldukları açısından yapılabileceği gibi önceden belirlenmiş kısıtlar eşik (sınır) değerleri çerçevesinde de yapılabilir.

6. İstatistiksel Mühendisliğin, Mühendislik Araştırmalarında Kullanımına İlişkin Uygulama Yöntemleri Üzerine Örnekler

6.1. İstatistik Mühendisliğinde Varyasyon-Standart Sapma Üzerine Temel Bilgi

Varyasyon~Değişim (variation) olan her ortamda bu değişimi azaltmak için istatistik yöntemlerine başvurulur. İstatistik Mühendisliğinde ise bu kavram "değişimi azaltmak" için sistematik bir algoritma" sunmayı önerir. Bu algoritma, endüstride seçilen süreç girdilerinin (input) ve çıktılarının (output) değerlerinin ölçülmesinin mümkün olduğu orta ölçekli üretim süreçlerine uyarlamaktır. Değişim (variation) kelimesi hem çıktının bir hedef değerinden sapması (deviation) hem de çıktının parçadan parçaya değişen değeri anlamında kullanılmaktadır. V6 pistonlarını üreten bir işlem

²V6 motor, tek bir krank çeviren, üçerli iki sıra halinde altı silindirden oluşan V tipi motordur. 6 Silindirli Motorlar, Motor bloğu üzerinde 6 adet silindirin olduğu yapıdır. Genellikle V şeklinde bir motor bloğu üzerinde üçerli şekilde silindirler sıralanarak 6 silindirli motor oluşturulur. Büyük motorlu araçlarda kullanılan 6 silindir yapısı, aracın ön kısmının da büyük olmasına neden olabilir.

sürecinde çap için hedef değeri 101.591 mm dir. Peş peşe üç pistonun mm türünden ölçülen çapları 101.592, 191.587, 101.598 olduğu saptanmıştır. Pistonların hiçbiri hedef çapına

sahip olmadığından ve hepsinin farklı çapları olduğundan her iki anlamda da değişim görülebilir. Problemler, azaltılmak istenen değişimin doğasını yakalayan uygun performans ölçümleri tanımlanarak formül tasarımına gidilecektir. Yüksek değişim kötü performansa, düşük müşteri memnuniyetine,

hurdaya çıkarılmaya ve yeniden işlemeye, karmaşık akıntı yönünde kontrol planlarına vb. yol açmaktadır. Eğer bu problemler çözülebilirse, giderler azaltılabilir ve kalite ve performans geliştirilebilir.



Şekil 1. Egzantirik mili

6.2. Algoritmanın Temeli Nasıl Oluşturulur?

Bir sürecin nasıl ve neden olduğu gibi davrandığının bilgisi artırılarak süreçte değişimi (variation) azaltacak uygun giderli değişikliklerin ortaya çıkartılacağı şeklindeki inançtır. Süreç bilgisini arttırmanın bir yolu deneysel olarak öğrenmek, yani gözlem ve deneyle öğrenmektir. İşte istatistik burada devreye girmektedir. Çünkü İstatistik deneysel olarak nasıl davranılacağını öğreten bir bilim koludur. İstatistik, “Süreç araştırması nasıl planlanır ya da planlanmalıdır?” ve “Eldeki veri nasıl yorumlanmalıdır?” gibi sorulara yanıtlar sağlar. Önerilen algoritma, istatistiksel yöntemlerin ve araçların var olan mühendislik bilgi ve teorisiyle birleştirilmesine bağlıdır. Algoritma kullanılarak, süreç davranışıyla ilgili öğrenmek için bir veya birden çok araştırma planlanacak ve gerçekleştirilecektir.

Çoğu kez, uygun maliyetli (giderli) olmaları için, önerilen değişimler temel tasarım değişiklikleri veya süreç donanımını değiştirmek yerine düzenli (ayarlı) bir duruma getirerek daha iyi bir süreç kontrolü veya değişiklikleri içerirler. Bu *düşük giderli değişimler genel değişimi azaltma yaklaşımlarının* bir kümesi içine sınıflandırılmış ve bir yaklaşımın erken göz önüne alınmasını zorlayacak *algoritma* yapılandırılır.

İşte istatistiksel mühendisliğin belirli amaçları bize aşağıdaki eylemler için yardımcı olacağı düşünülür:

- Uygun giderli *değişim azaltmanın* nasıl başarılacağıyla ilgili stratejik olarak düşünmeyi sağlamak.

- *Değişim kaynaklarını ve bunların süreç gelişimindeki rolünü anlamak.*
- *Deneysel yöntemlerin daha iyi nasıl kullanılacağını; yani, bir süreç araştırmasının sonuçlarını planlamanın, uygulamanın ve analiz etmenin etkili ve verimli yollarını öğrenmek.*

Ana amaç önerilen algoritma kullanılarak dile getirilebilen problem türlerinin örneklerini ortaya çıkartmaktır. Yani problemler tartışılmaya açık, fakat çözüm yolları devrede olmayacaktır. Aşağıda verilecek örnekle algoritma, yaklaşım ve yöntemleri uygulayarak kazanımlar nasıl elde edilir onları göreceğiz.

6.3. Motor Mühendisliğinden Bir Örnek: Motor Bloğu Sızıntıları Problemi ve İstatistiksel Çözüm Önerisi

Şekil 1.2’ de gösterildiği gibi, bir eksantrik milinin lobunun geometrisi bir motorun işleminde kritiktir. Eksantrik mili loblarının dönmesi motor vanalarının açılmasını ve kapanmasını yönlendirmektedir. Görüntülenen eksantrik milinin üçü beyaz oklarla gösterilen 12 lobu vardır.

Yandan görüldüğünde, lobun tabanı (merkez çizgisinden 60°) ideal olarak bir çemberin yayıdır. Şekil 1.3 bir lobun tabanı boyunca ideal, çemberden (mm olarak) sapmanın bir izidir.

Lob geometrisiyle ilişkili ölçülen altı kritik karakteristikten *taban çemberi salgısı* tarihsel olarak en sorunlu olanı ve bu yüzden bir *değişim azaltma* alıştırmasının odağı olarak seçilmişti. Taban çemberi (TÇ) salgısı gerçek lob geometrisinin ideal çemberden (60° lik yay üzerinde en fazla – en az) sapmasının pozitif bir ölçümüdür. 0.000 mm lik bir deger tabanın tamamen dairesel olduğu anlamına gelmektedir.

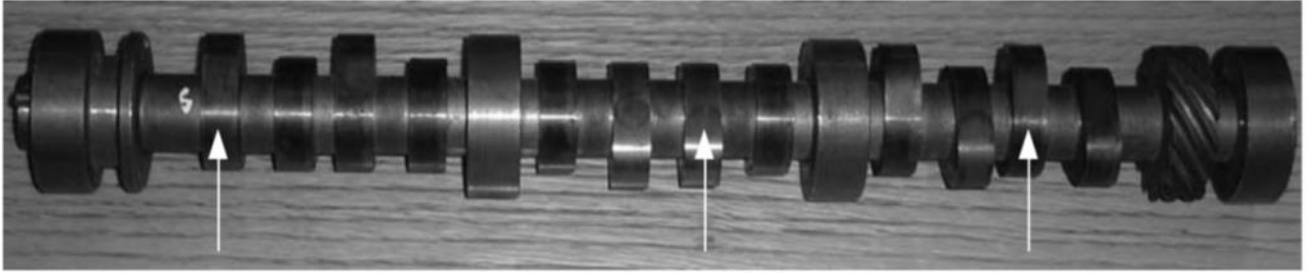
İzin verilebilir en fazla TÇ salgısı 0.040 mm veya 40 mikrondur.

12 gün boyunca günde 9 parça olmak üzere 108 parçalık bir örnek toplanmıştır ve her eksantrik mili üzerinde 12 lobun her biri için TÇ salgısı ölçülmüştür. (mikron cinsinden kaydedilmiş) 1296 salgı ölçümü ve lobların diğer bazı geometrik karakteristikleri camshaft lobe runout baseline (eksantrik mili lob salgısı taban çizgisi) dosyasında mevcuttur. Tüm loblar üzerindeki salgı değerlerinin bir histogramı Şekil 1.4 da gösterilmiştir.

Tüm loblar için TÇ salgısı şartname sınırının çok altındaydı ama bu kritik karakteristiğin motor performansı üzerindeki etkileri nedeniyle, yönetim süreci geliştirmek için bir proje başlatılmıştır. Burada amaç süreci histogramın sağ kuyruğu sola

kayacak ve lobların daha büyük bir miktarı sifıra yakın salgiya sahip olacak şekilde değiştirmekti. Salgının pozitif olması gerektiğinden, süreci tob loblar üzerindeki alt salgiya göre ayarlamak mümkün değildi. Bu bağlamda, değişim azaltılması yapılır ki bunun anlamı salgıları hedef değeri olan 0 mikrona yakın olan daha yüksek miktarda lob olacağıdır.

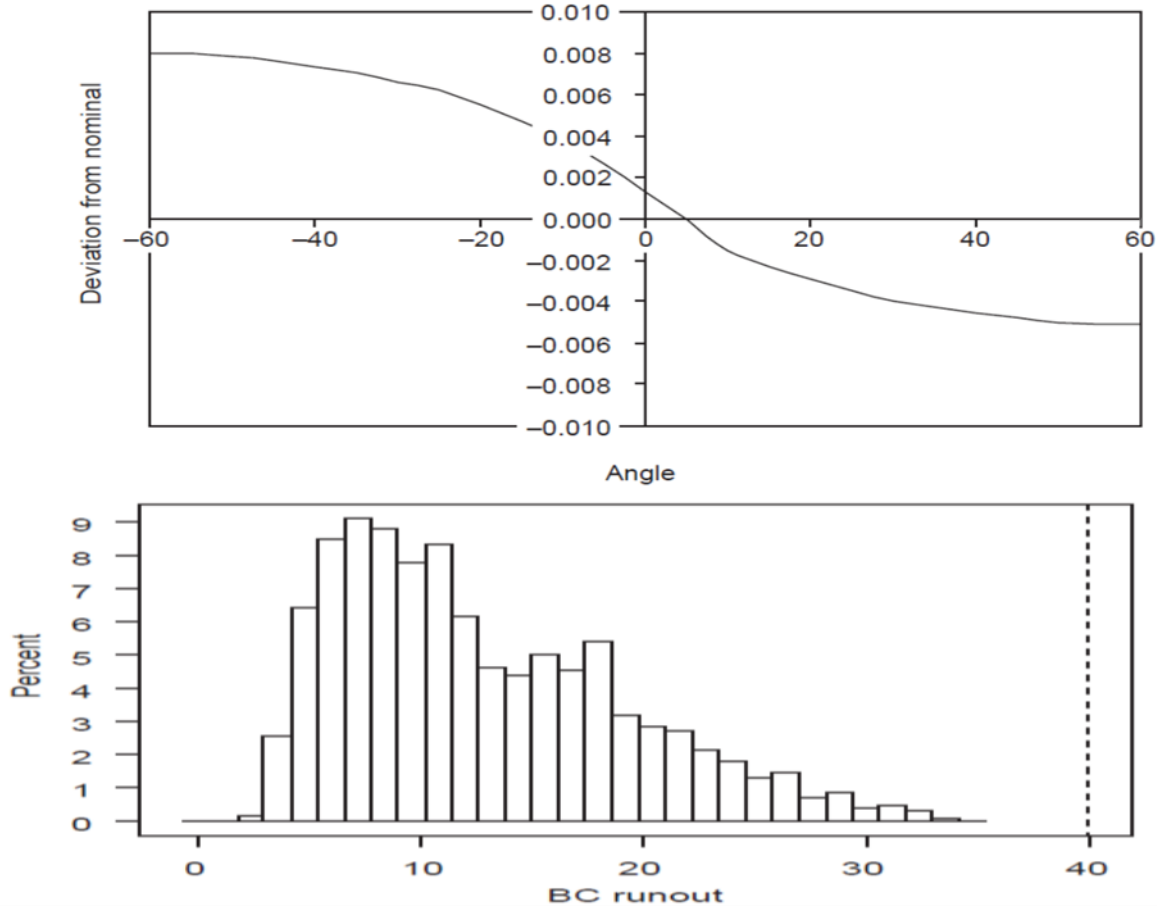
Tek yönlü bir şartnamesi ve fiziksel bir alt sınırı olan diğer birçok süreç karakteristiği vardır. Örnekler arasında pürüzsüzlük, gözeneklilik, koniklik v.b. vardır. Bu tür karakteristikler cinsinden tanımlanmış süreç performansı hedef değeri civarındaki değerlerin yoğunluğu artırılarak geliştirilebilir. Yani, hedef değeriyle ilgili süreç çıktısındaki değişim azaltılır.



Şekil 2. Eksantrik mili (Eksantrik mili de motor ile bağlantılı olarak işlem yapan bir araç içi parçadır. Araç içlerinde yer alan supapların gerekli zamanda açılıp kapanmasını sağlayan ve üzerinde kam bulunan küresel demir ya da çelik alaşımdan yapılan bir sistemdir) [18].



Şekil 3. İdealden lob sapması



Şekil 4. Eksantrik mili lob TÇ salgısının histogramı (kesikli çizgi şartname sınırını vermektedir).

7. Sonuç

Süreç çıktılarında değişimi azaltmak süreç gelişiminin kilit bir parçasıdır. Süreç değişimini azaltmak için İstatistiksel Mühendislik olarak adlandırılan bir algoritmanın etrafında bir yapılandırma gerekliliğinin önemine vurgu yaptık. Algoritmalar örneğin üretim ve montaj süreçlerinde mevcut yüksek ile orta ölçekli kronik problemleri çözmek için tasarlanırlar. Algoritmanın ana temeli eğer bir sürecin nasıl ve neden yaptığı gibi davranışıyla ilgili bilgi arttırılırsa değişimi azaltacak süreçte uygun maliyetli değişiklikler keşfedileceği şeklindeki bir inanç olduğuna inanıyoruz.

Süreç bilgisini azaltmanın kilit bir yolunu açıklarsak deneysel olarak-yani, gözlem ve deneyle-öğrenmektir. Biz de çalışmamızda Deneysel araştırmaları planlamak ve analiz etmek için kısaca SPVAS (Soru, Plan, Veri, Analiz, Sonuç) olarak bilinen bir çerçeveyi detaylı biçimde kullandık gerek dil konusunda gerekse mühendislik deneysel araştırmalarda nasıl kullanıldığına ilişkin uygulamalardan söz ettik. Çünkü SPVAS çerçevesi İstatistiksel Mühendislik algoritmasının birçok aşamasında uygun araştırmaların sonuçlarını planlamaya, analiz etmeye ve yorumlamaya yardımcı olmak için kullanımının çok önemli olduğunu deneysel problemlerle gösterdik. Hangi veri olursa olsun mutlaka bir araştırma için amaç ve planlamanın bir algoritmanın evresinin, daha önceki araştırmalardan ortaya çıkmış sonuçlarla ilgili bilgilerin, eğer konu mühendislik ise süreç ve bilgilerinin varlığına bağlı olduğunun sonucuna ulaşmış olduk. Bu çalışmayla araştırmacıların yaklaşımların ve yöntemlerin kullanımında güven kazanmasına yardımcı olması için araştırmaları cesurca

denemesi ve veri kümelerini (hatta büyük veri ~ big data) keşfetmesini tavsiye ediyoruz.

Yazar Katkısı

Veri iyileştirme -Yazar Adı Soyadı(AS); Biçimsel analiz - AS; Araştırma - AS; Deneysel performans - AS; Veri toplama - AS; Veri işleme - AS; Literatür taraması - AS; Yazan - AS; İnceleme ve düzenleme - AS...

Çıkar Çatışması Beyanı

Yazarlar, bu makalenin araştırılması, yazarlığı ve/veya yayınlanması ile ilgili olarak herhangi bir çıkar çatışması beyan etmemiştir. İkar Çatışması Beyanı

Kaynaklar

- [1] Mazmanoğlu, Adnan. (2016) Herkes için İstatistik Analiz Yöntemleri I, 1.Baskı, Ankara, Nobel Akademik Yayınları
- [2] Grisebach, Manon Maren. (1995). Edebiyat Bilim'inin Yöntemleri, (Çev.: Arif Ünal), Ankara: Atatürk Kültür, Dil ve Tarih Yüksek Kurumu, Atatürk Kültür Merkezi Yayını Sayı 99
- [3]https://www.google.com/search?q=g.u.yule+ve+c.b.+williams&tbm=isch&ved=2ahUKEwjjjeDSuur_AhUlgv0HHZJa_CiUQ2-cCegQIABAA&oq=g.u.yule+ve+C.B.+Will&gs_lcp=CgNpbWcQARgA
- [4] https://de.wikipedia.org/wiki/Wilhelm_Fucks [10] M. G. Kendall Yule, G. Udney, An Introduction to the Theory of Statistics. Fourteenth edition, revised and enlarged Hardcover

– 1 Jan. 1965

- [6] https://en.wikipedia.org/wiki/Christian_Morgenstern
- [7] <https://www.poetryfoundation.org/poets/august-stramm>
- [8] <https://tr.wikipedia.org/wiki/Dadaizm>
- [9] <https://www.jstor.org/stable/455407>
- [10] [https://www.themaggar.com/gertrude-stein-kimdir-yasami/\(Who is gertrude-stein\)](https://www.themaggar.com/gertrude-stein-kimdir-yasami/(Who%20is%20gertrude-stein))
- [11] https://de.wikipedia.org/wiki/Helmut_Hei%C3%9Fenb%C3%BCtte
- [12] C. E. SHANNON, A Mathematical Theory of Communication, Reprinted with corrections from The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, July, October, 1948.
- [13] Musođlu, Erdal, (2016-Ekim). Claude Shannon, sayısal

devrimi bařlatan mhendis, İstanbul: HBT sayı 28-7 Ekim

- [14] Sezgin, Fatin. (1993) Dil ve Edebiyatta İstatistik ve Bilgisayar Uygulamaları, 1. Basım, İstanbul: Dergâh Yayınları
- [15] İstatistik Metot Eleřtiri Kuramı Teorisi Hakkında Kısaca Özet Bilgi, <https://www.edebiyatfakultesi.com/istatistik-metot-elestiri-kurami.htm>, 06.06.2023
- [16] P.Berkhin , A Survey of Clustering Data Mining Techniques, https://link.springer.com/chapter/10.1007/3-540-28349-8_2
- [17] Cleaver, G.B., Nanopoulos, D. V. A. E. FARAGGI, and T. TER VELDHUIS, TOWARDS STRING PREDICTIONS, International Journal of Modern Physics AVol. 16, No. 21, pp. 3565-3581(2001)
<https://doi.org/10.1142/S0217751X01004918>
- [18] <https://otolye.com/blog/eksantrik-mili>