Journal of Happiness Studies (2025) 26:39
https://doi.org/10.1007/s10902-025-00881-9

RESEARCH PAPER

Development and Validation of a Real-Time Happiness Index Using Google Trends™

Talita Greyling[1,2,3] · **Stephanié Rossouw**[1,2,3]

Accepted: 16 February 2025

Abstract

It is well-established that a positive relationship exists between happiness and the economic outcomes of a country. Traditionally, surveys have been the main method for measuring happiness, but they face challenges such as "survey fatigue", high costs, time delays, and the fluctuating nature of happiness. Addressing these challenges of survey data, Big Data from sources like Google Trends™ and social media is now being used to complement surveys and provide policymakers with more timely insights into well-being. In recent years, Google Trends™ data has been leveraged to discern trends in mental health, including anxiety and loneliness, and construct robust predictors of subjective well-being composite categories. We aim to construct the first comprehensive, near real-time measure of population-level happiness using information-seeking query data extracted continuously using Google Trends™. We use a basket of English-language emotion words suggested to capture positive and negative affect and apply machine learning algorithms—XGBoost and ElasticNet—to identify the most important words and their weight in estimating happiness. We demonstrate our methodology using data from the United Kingdom and test its cross-country applicability in the Netherlands by translating the emotion words into Dutch. Lastly, we improve the fit for the Netherlands by incorporating country-specific emotion words. Evaluating the accuracy of our estimated happiness in countries against survey data, we find a very good fit with very low error metrics. Adding country-specific words improves the fit statistics. Our suggested innovative methodology demonstrates that emotion words extracted from Google Trends™ can accurately estimate a country's level of happiness.

Keywords Happiness · Google trends™ · Big data · XGBoost · Machine learning

JEL Classification C53 · C55 · I31

✉ Stephanié Rossouw
stephanie.rossouw@aut.ac.nz

[1] School of Economics, University of Johannesburg, Johannesburg, South Africa

[2] School of Social Sciences and Humanities, Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand

[3] Research Fellow With the Global Labor Organization, Essen, Germany

Published online: 26 February 2025

1 Introduction

Measuring well-being using subjective measures is essential since it is accepted that a positive relationship exists between happiness and the economic outcomes of a country. People's happiness profoundly affects these outcomes, including productivity, labour market performance, and future income (Bryson et al., 2016; Piekalkiewicz, 2017). Increased happiness also positively affects a nation's social and health sectors (Kim et al., 2015), fosters altruistic behaviour and enhances various cognitive and social capabilities (Kasser & Ryan, 1996; Williams & Shiaw, 1999). Happier individuals are healthier, live longer, and generally report higher levels of life satisfaction. They are more likely to avoid high-risk activities and take preventive measures to reduce potential risks.

Traditionally, the primary source for measuring people's happiness has been survey data. However, in a post-pandemic era, people experience 'survey fatigue'. Moreover, conducting surveys is expensive and often results in data that is delayed by up to two years, which may also be affected by non-response bias (Callegaro & Yang, 2018; Rossouw & Greyling, 2020).

To overcome these limitations of survey data, researchers have turned to Big Data to measure and track people's happiness. Measuring people's happiness using Big Data adds an additional benefit since decision-makers are often confronted with short-term horizons and imperfect information. Therefore, they need an immediate source of information regarding a country's mood so that people's needs and concerns guide policies for achieving collective outcomes (Rossouw & Greyling, 2024). Real-time information from Big Data will also allow decision-makers to gauge possible reactions to the proposed legislature to mitigate potentially violent and destructive outcomes (Greyling & Rossouw, 2022). Notably, the work done by Dodd and Danforth (2010), Iacus et al., (2015 and 2022), and Greyling and Rossouw (2019) set the way to harness the power of Big Data. All three studies utilised Twitter data to construct happiness or subjective well-being measures (see Sect. 2.2 for full discussion).

Unfortunately, with Elon Musk's purchase of Twitter (now X), all academic licenses were suspended, and access to Twitter data was stopped, effectively closing the book on academic research. Therefore, researchers focusing on measuring happiness in real time had to resort to other Big Data sources.

Recent research has demonstrated the value of information-seeking query data in forecasting social phenomena. Carammia et al. (2022) utilised a Dynamic Elastic Net (DynENet) model to predict asylum-related migration flows by integrating administrative data with non-traditional sources such as internet searches and geolocated event data. Their work has underscored the broader applicability of internet-derived data for real-time social monitoring.

Therefore, we aim to explore an innovative methodology to accurately estimate happiness levels and their evolution at a country level from information-seeking query data based on a carefully curated selection of English emotion words extracted continuously from Google Trends™. In our proof of concept, we validate our index using the happiness survey measure from the United Kingdom's (UK) Office for National Statistics (ONS) (referred to as True Happiness). Our second aim is to explore whether the same selected basket of English words translated into a different language (Dutch) with the same weights can also successfully estimate happiness in the Netherlands. Here, we validate our equation against the Dutch Time Use data. Our last aim is to follow the same methodology for our initial derived UK happiness equation using the Dutch Time Use

happiness measure as the outcome variable for predictions. Here, we use the initial basket of emotion words and add country-specific words to attain a more accurate estimate of happiness in a country.

Previous studies (see Sect. 2.3 for full discussion) leveraging Google Trends™ data measured trends in mental health, including depression, anxiety, and loneliness (Brodeur et al., 2020; Foa et al., 2022; Ford et al., 2018), constructed robust predictors of subjective well-being composite categories in the United States (Algan et al., 2019) and nowcasted national average subjective well-being (Murtin & Salomon-Ermel, 2024). However, none of these studies attempted to estimate county-level happiness measured and updated in almost real-time. Therefore, to our knowledge, our near real-time measure of population-level happiness using information-seeking query data extracted continuously using Google Trends™ in countries is a first of its kind.

To construct our happiness index, we start by identifying emotion words that are grounded in the theoretical framework of the works of Watson et al. (1988), Thompson (2007), and Diener et al. (2010). We use a basket of 69 English-extracted emotion words suggested to capture affect in the delivery of Positive and Negative Affect Schedules Extended (PANAS-X), International Positive and Negative Affect Schedule Short-Form (I-PANAS-SF), Scale of Positive and Negative Experience (SPANE), and various other studies including Engelen et al. (2006), Kahn et al. (2007), Dodd and Danforth (2010), Ford et al. (2018), Algan et al. (2019) and Boyd et al. (2022).

After selecting the abovementioned 69 words, we refined the list for the UK by testing the correlation of each word with True Happiness and retained only those that showed a statistically significant correlation. To further narrow the selection, we applied eXtreme Gradient Boosting (XGBoost), a machine learning algorithm, to rank the words based on their "gains," indicating the most important predictors of True Happiness. Next, we determine the weighting of the words (features) using the estimated coefficients of each word derived from predicting True Happiness using ElasticNet, a machine learning regression algorithm.

To test the accuracy of our derived equation to estimate UK happiness, we compare it to the UK's True Happiness measure. The results show a good fit (RMSE = 0.09), indicating that our index has a high level of accuracy in estimating happiness at the country level.

We perform various additional robustness tests using different frequencies of the extracted data (time-invariance) and unseen datasets (other periods). We also validate our Google Trends™ happiness index against another Big Data measure, namely the World Health Organisation's Early AI-supported Response with Social Listening (EARS). The dataset does not measure happiness but mental health and loneliness; therefore, a significant negative correlation will indicate our index's robustness.

To address our second research question, we use the same derived equation containing the same selected words and weights for the UK translated into Dutch to predict happiness. After applying the equation and estimating happiness in the Netherlands for 2011 and 2020, we evaluated the fit against the happiness measure from the Dutch Time Use survey data. We find different results considering different time periods. The results show a good fit for 2011 (RMSE = 0.08) and a weaker fit for 2020 data (RMSE = 0.43). However, the Time Use data quality weakened over time with fewer respondents and fewer observations, which might have contributed to the weaker fit.

To address our last research question, we re-estimate our happiness index, including country-specific Dutch words and validate it using Dutch Time Use data. The error metrics (RMSE = 0.05) for 2011 indicate a marginally better fit than using the primary selected emotion words and their weights and a significantly better fit for 2020. We find an overlap

of the most important words in the UK and the Netherlands, but adding a few country-specific words improves the fit statistics.

Therefore, our results show that we achieve an acceptable fit using the same basket of words across countries, demonstrating our model's adaptability and scalability to different cultural and linguistic contexts. However, we can improve the fit by including country-specific emotion words to accurately estimate happiness levels from information-seeking query data extracted continuously from Google Trends™.

By achieving our aim of developing and validating a real-time happiness index, we offer governments and other stakeholders access to timely and relevant information about the mood of their citizens, which is applicable for decisive decision-making at significantly lower costs than survey data with a possibility to automate, to some extent, the process of measuring happiness.

The rest of the paper is structured as follows. The next section contains our theoretical framework for the emotion words and provides a literature review pertaining to studies that measured real-time happiness or subjective well-being. The data, selected variables and methodology are discussed in Sect. 3. The results follow in Sect. 4, while the paper concludes in Sect. 5.

2 Literature Review

This section first discusses the theoretical framework for our emotion words and studies that explored the use of affect words. The next section discusses studies that developed real-time measures for happiness or subjective well-being using social media or search engines.

2.1 Measuring Affect

2.1.1 Theoretical Framework

Watson et al. (1988) developed the Positive and Negative Affect Schedules (PANAS). The original PANAS included a 20-item bidimensional scale, which were broadly independent and discrete dimensions of affect rather than polar opposites on a continuum. These words were "Interested", "Distressed", "Excited", "Upset", "Strong", "Guilty", "Scared", "Hostile", "Enthusiastic", "Proud", "Irritable", "Alert", "Ashamed", "Inspired", "Nervous", "Determined", "Attentive", "Jittery", "Active" and "Afraid".

Watson and Clark (1994) expanded the original PANAS, known as PANAS-X, by creating a 60-item measure, which now expanded the two original higher-order scales to include 11 specific affects: "Fear", "Sadness", "Guilt", "Hostility", "Shyness", "Fatigue", "Surprise", "Joviality", "Self-Assurance", "Attentiveness", and "Serenity". The PANAS-X, therefore, measures mood at two different levels.

The original set of 20 words in the PANAS has faced some criticism. Validation studies using structural equation modelling, such as those by Crawford and Henry (2004), indicate that the most accurate models emerge when correlations are allowed between errors of items within the same word clusters from which the PANAS was initially developed (refer to Zevon & Tellegen, 1982, for word-cluster descriptors). These item covariances suggest a high degree of redundancy among certain PANAS items with similar meanings. Crawford and Henry's (2004) analysis demonstrated that the 10 items of the Negative Affect (NA)

scale form five pairs with significant covariance: "distressed" and "upset," "guilty" and "ashamed," "scared" and "afraid," "nervous" and "jittery," and "hostile" and "irritable." Similarly, the Positive Affect (PA) scale's ten items cluster into four groups with shared variance. Two groups contain three items each: "interested," "alert," and "attentive," and "excited," "enthusiastic," and "inspired." The remaining two groups are formed by two pairs: "proud" and "determined," and "strong" and "active." These findings suggest that reducing the number of PANAS items may be possible without significantly compromising the PA and NA scales' content coverage or internal consistency.

Kercher (1992) created a shortened version of the original PANAS, reducing it to 10 items: "Excited," "Enthusiastic," "Alert," "Inspired," "Determined," "Distressed," "Upset," "Scared," "Nervous," and "Afraid." However, Mackinnon et al. (1999) noted that Kercher's abbreviated version included items with high covariance, which undermines content validity while artificially increasing internal consistency reliability. Furthermore, the full PANAS and Kercher's (1992) shortened version contained items with unclear or ambiguous meanings to both native and non-native English speakers from outside North America. For instance, many non-native English speakers do not understand the term "jittery," which is considered colloquial in many dictionaries. Additionally, Mackinnon et al. (1999) found that even among native English speakers, the item "excited" in Kercher's short form correlated significantly with both Positive Affect (PA) and Negative Affect (NA), suggesting it carries dual meanings for some.

To address redundancy issues and ambiguous meanings in different research contexts, Thompson (2007) developed the I-PANAS-SF (International Positive and Negative Affect Schedule Short Form). This new version was validated across various national, cultural, and occupational groups, demonstrating strong psychometric properties, including cross-sample stability, internal consistency, temporal stability, cross-cultural factorial invariance, and convergent and criterion-related validity. The I-PANAS-SF uses the question stem "Thinking about yourself and how you normally feel, to what extent do you generally feel:" and includes 5 Positive Affect and 5 Negative Affect items: "Upset," "Hostile," "Alert," "Ashamed," "Inspired," "Nervous," "Determined," "Attentive," "Afraid," and "Active."

2.1.2 Studies Focusing on Affect Words

A study applicable to the aim of this paper is Jovanović et al. (2022), who used the SPANE (Scale of Positive and Negative Experience), created by Diener et al. (2010), to determine its cross-cultural utility by measuring the invariance of the SPANE. The SPANE consists of 12 items designed to assess how often positive (SPANE-P subscale) and negative (SPANE-N subscale) emotions are experienced. It was created to address the limitations and challenges identified in previous emotion measurement tools, such as the PANAS. These include "Positive", "Negative", "Good", "Bad", "Pleasant", "Unpleasant", "Happy", "Sad", "Afraid", "Joyful", "Angry" and "Contented". Jovanović et al. (2022) focused on 13 countries: the United States, Turkey, Spain, Serbia, Portugal, Poland, Japan, Italy, India, Greece, Germany, Colombia and China. They found that SPANE's positive emotion terms, "positive", "good", "pleasant", "happy", "joyful", "contented", and general negative emotion terms, "negative" and "unpleasant", could be suitable for studies on emotions and well-being in a cross-cultural project.

Apart from the above, we also relied on the LIWC-22 dictionary (Boyd et al., 2022), a text analysis tool designed to assess language's psychological, social, and linguistic dimensions. LIWC-22 builds on previous versions by expanding its dictionary, refining its

algorithms, and enhancing its usability for psychology, linguistics, and other social sciences researchers. LIWC was also validated by Kahn et al. (2007) as a valid tool for measuring emotional expression.

Other studies we used to identify emotion words include Engelen et al. (2006), which validated the Dutch versions of the PANAS, confirming their reliability and applicability for Dutch-speaking populations. The authors emphasised the importance of cultural adaptation when translating psychological measures. While the PANAS was originally developed in English, the study found that culturally sensitive translations retain the measure's effectiveness and ensure it remains meaningful across different linguistic and cultural contexts. We also considered the study done by Banerjee (2018), who used Google search data to study the patterns in public interest and concern related to the "internet", "anxiety", and "happiness", exploring how they are interrelated and vary across different countries and cultures. The author found that search volume data indicate significant interest in understanding how these topics connect to daily life, personal well-being, and mental health and that searches for "internet" often correlate with searches for "anxiety" and "happiness." This suggests a potential link between internet use and psychological states, where people might be using the internet both as a tool for coping with anxiety and as a means to seek or understand happiness.

Our last three studies, Dodds and Danforth (2010), Algan et al. (2019) and Ford et al. (2018), are discussed in Sects. 2.2 and 2.3 since they relied on measuring happiness or subjective well-being using Twitter or Google Trends™.

2.2 Measuring Happiness or Subjective Well-Being Using Twitter

The pioneering research conducted by Dodd and Danforth (2010), Iacus et al., (2015, 2022), and Greyling and Rossouw (2019) are essential for measuring subjective well-being or happiness using Big Data sources like Twitter and will be further discussed below.

The Hedonometer was one of the pioneering tools developed to measure happiness in almost real-time using Big Data. Initiated by Dodds and Danforth (2010) at the end of 2008, the project tracked daily happiness levels, creating a continuous time series from late 2008 to May 2023 when Twitter (now X) suspended all academic licenses (refer to Dodds et al. (2011) for the foundational study). To begin, the authors merged the 5,000 most common words from four sources: Twitter posts, articles from the New York Times, Google Books, and Music lyrics. After merging the words, they are left with a composite set of around 10,000 unique words. They then used Amazon's Mechanical Turk to rate each word's happiness on a scale from 0 (unhappy) to 10 (happy), with "laughter" scoring the highest at 8.5 and "terrorist" scoring the lowest at 1.3. To construct the Hedonometer, they bin all the tweets extracted daily; however, only words recognised as English were included. The bin includes, on average, 200 million words extracted worldwide daily. Using a bag-of-words methodology, they assign a happiness score to each word, which is then averaged to produce a daily happiness index.

Iacus et al., (2015, 2022) were among the first to create a composite index of subjective and perceived well-being, encompassing various aspects of both individual and collective life. However, the measure was developed based on a priori-defined dataset without real-time predictive power. They developed their Subjective Well-being Index (SWBI) by applying an Integrated Sentiment Analysis (iSA)to tweets from Italy (starting in 2012) and Japan (beginning in 2015). The SWBI consists of eight components that reflect three distinct areas of well-being: social well-being, personal well-being, and well-being at work,

with the final score being the average of these components. For instance 2015, Italy's SWBI averaged 48.7, while Japan's averaged 54.4. Carpi et al. (2022) used Random Forest and ElasticNet to analyse the impact of external factors such as data on the spread of COVID-19, economic indicators, air quality, internet searches and mobility data on the SWBI for Japan and Italy. Among others, they found that data on the spread of COVID-19, such as the number of deaths and cases, were more important for the SWBI in Japan than in Italy. Air quality was only relevant to the SWBI in Italy, whereas economic indicators were more relevant to the SWBI in Japan.

The *Gross National Happiness.today* project was launched by Greyling and Rossouw (2019) to determine national happiness levels (evaluative mood) in near real-time during different social, economic, and political events. They created their high-frequency daily time-series data by extracting live tweets and applying natural language processing (NLP) to analyse the sentiment. The sentiment analysis uses a lexicon-based approach incorporating tools like TextBlob, VADER, Sentiment140, and NRC, classifying tweets as positive, negative, or neutral. A balancing formula calculates a happiness score, which is averaged hourly and daily to provide near real-time time-series data. The scores range from 0 to 10, with 5 representing a neutral state, neither happy nor unhappy. In 2020, the project expanded to measure eight distinct emotions based on Plutchik's (1980) wheel of emotions, generating daily time-series data for each emotion. This project was also temporarily halted after Twitter (now X) suspended all academic licenses.

2.3 Measuring Mental Health, Life Satisfaction and Subjective Well-Being Using Google Trends™

Murtin and Salomon-Ermel (2024) used Google Trends™ data to nowcast national average subjective well-being estimates for 38 OECD countries since 2010. To train their nowcasting models, they collected a large sample of time series from Google Trends™, covering 158 topics and 914 categories of searches chosen based on their relevance according to the American Time Use Survey, the OECD Well-being framework, and the domains of life satisfaction in happiness studies. The authors created a condensed version of the Google Trends™ dataset, where multiple time series were aggregated into subtopics based on the facets of the OECD Better Life Index. They derived 42 composite variables representing different dimensions of well-being. Their control variables included GDP per capita (with constant prices and purchasing power parity), the inflation rate, and the participation rate for individuals between 15 and 64 years old. They utilised large, customised micro-databases to improve model training on thoroughly pre-processed Google Trends™ data. Their findings related to life satisfaction indicated that the most accurate one-year-ahead predictions were achieved using a meta-learning approach that integrates forecasts from an ElasticNet model (both with and without interactions), a Gradient-Boosted Tree, and a Multi-layer Perceptron. Consequently, for 38 countries from 2010 to 2020, the out-of-sample prediction of average subjective well-being achieved an R^2 of 0.830.

Algan et al.'s (2019) study investigated how changes in internet search volumes can model and estimate subjective well-being in the United States. The authors used data from Google Trends™ to analyse the relationship between search behaviours and well-being measures from Gallup Analytics, covering the period from 2008 to 2013. The study developed national and state-level models using search data condensed into composite categories (e.g., job search, civic engagement, healthy habits) that reflect different life dimensions. Both models showed high out-of-sample predictive accuracy and effectively

captured well-being trends. Using stepwise regression, they found that searches related to job search, civic engagement, and healthy habits consistently predict well-being (Gallup's indicators for "life evaluation today", "life evaluation in 5 years", "happiness", "laugh", "learn" and "respect") across multiple datasets and models. Job search terms are generally associated with lower well-being, while searches about civic engagement and healthy habits correlate with higher well-being.

Brodeur et al. (2020) utilised Google Trends™ data to examine the impact of government-imposed lockdowns on mental health and well-being. Their findings revealed a negative effect, indicated by increased searches related to sadness, worry, and loneliness. Foa et al. (2022) used two years (2020–2021) of Google Trends™ data from six English-speaking countries, along with weekly data from YouGov's Great Britain Mood Tracker Poll, to explore changes in subjective well-being throughout the COVID-19 pandemic. Using Google search terms such as "stress", "boredom", "frustration", "sadness", "loneliness", "feeling scared" ("fear"), "apathy", "happiness", "contentment", "energy", "inspiration" ("artistic inspiration"), and "optimism", they found that across the population, a decrease in affect tend to be associated with pandemic outbreaks. Furthermore, they found that while negative affect increased at the onset of lockdown, countries typically revert to baseline levels within three weeks at most, after which a net decrease in negative affect is observed.

In their study, Ford et al. (2018) used 15 terms from the PANAS-X to test the extent to which aggregated scores of emotion-related Google search queries are valid as indicators of subjective well-being at the US state and metro area levels. The selected terms included "afraid", "anxiety", "depression", "fatigue", "fear", "lonely", "nervous", "scared", "sleepy", "stress", "tired", "energetic", "enthusiastic", "happy", and "strong". The authors examined correlations between Google search scores and Gallup-Healthways measures of experienced negative emotions, namely "stress", "worry", "anger", and "sadness", as well as a composite measure combining these four emotions. They found that "afraid" was the most robust search term as it had significant associations with all the Gallup-Heatlthways indicators. Searches for "fear" were positively related to Gallup's "stress", "worry", and general negative affect. Searches for "scared", "lonely", and "nervous" were related to Gallup indicators of "anger" and "sadness", although "nervous" was also related to general negative affect. Interestingly, search scores for "depression" and "stress" were negatively related to Gallup "anger", while search scores for "anxiety" did not correlate with any Gallup items. The searches related to low arousal, e.g., "tired" and "fatigue", showed no relationships with Gallup indicators.

3 Data and Methodology

3.1 Data

3.1.1 Primary Dataset – Big Data Using Google Trends™

Google Trends ™ is an open data service provided by Google Inc., allowing researchers to explore the temporal patterns of internet search activity based on specific keywords. It offers access to a single metric: the Relative Search Volume (RSV), a standardised measure reflecting search activity relative to the chosen time frame and geographic region. The RSV values range from 0 to 100, enabling comparisons of search volume trends across different queries, time periods, and locations (Houghton et al., 2023). Data

from Google Trends™ excludes certain data from searches. First, it excludes topics of interest where the interest is very low. Google Trends™ only analyses data for popular terms, so search terms with low volume appear as 0 for a given time period. Second, Google Trends™ excludes duplicate searches. It removes repeated searches by the same user within a short time frame to enhance overall accuracy. Lastly, Google Trends™ excludes special characters by filtering out queries with apostrophes and other special characters.

Working with Google Trends ™ data has certain limitations. First, Google Trends™ data loses predictive power over time due to changes in search activity and the interface of Google Search itself (for example, auto-suggestion). For our purposes, this means that we need to regularly review the affect words in Table 1 and make any necessary adjustments to the weighting within our regression models. This will also allow us to periodically update and revise our index, which is also important for the model to incorporate social changes that could necessitate a re-weighting of the components.

Second, the search volume value on any given day cannot be directly compared across different terms because each term is normalised to its own maximum value. To resolve this issue, we standardise all search volumes to have a mean of zero and a standard deviation of one, focusing on changes in volume within each search term rather than relative differences between terms.

Third, Google Trends™ data presents estimation challenges because it does not provide raw search volumes; instead, it represents the proportion of total searches containing a specific keyword over a given period, normalised so that the highest value is 100. This normalisation affects the data interpretation in two main ways: first, the values directly obtained from Google Trends™ can be complex to interpret since they are influenced by both the search volume of the keyword and the overall search activity. Second, values on a given day cannot be compared between different terms, as each is scaled to its maximum. We standardised all search volumes to focus on within-term changes to address these issues rather than comparing absolute search volumes.

Fourth, while Google Trends™ provides a valuable real-time measure of public interest, its data is inherently influenced by demographic biases in search behaviour. Internet access, digital literacy, and platform preferences vary significantly across age groups, socioeconomic statuses, and geographic regions, leading to the potential overrepresentation of certain populations. This is a limitation in general when using digital data sources; therefore, any data derived from these sources must be validated against

Table 1 The 69 words extracted to establish those with the highest correlation with True Happiness

Great	Joke	Attentive	Cry	Punish	Wellbeing	Angry
Party	Joy	Inspired	Dead	Reject	Well-being	Cancer
Game	Love	Active	Depressed	Sad	Suicide	Divorce
Comedy	Music	Alone	Disease	Sick	Sleep	Hopeless
Friendship	Pleasure	Abuse	Fear	Stress	Sadness	Pain
Fun	Win	Afraid	Hate	Tired	Boredom	Weak
Good	Movie	Anxiety	Headache	Worry	Depression	Joyful
Happy	Song	Anxious	Kill	Wrong	Loneliness	Contented
Health	Friend	Bad	Lonely	Panic	Ashamed	Determined
Hope	Alert	Crime	Nervous	Upset	Unpleasant	

survey data, which is not susceptible to this limitation. In the current study, we validate our measures against survey measures of happiness.

Additionally, relying on Google Trends™ data to measure real-time happiness presents a risk over which researchers have no control, primarily due to the unpredictability of large tech platforms in maintaining their services. Previously, major players like Google, Meta and Twitter (now X) have abruptly discontinued services (e.g., Google Mobility Maps), restricted API access (e.g., Instagram, which impacted numerous applications, including popular dating apps like Tinder and Hinge, as well as the journaling app Day One), modified algorithms without warning (as frequently seen with Meta's advertising platform) or shut down its API service altogether (Twitter). Such changes can disrupt research or analytics that depend on consistent data streams. Therefore, we continuously explore alternative digital resources to derive well-being measures.

Apart from the above, it is important to consider that there may be a disconnect between survey-based measures and online search behaviours, which should be considered when selecting and using survey-derived keywords in the context of internet search data. This implies that if we consider positive and negative emotions, we must also consider the search actions that will be taken if these emotions are experienced, thereby necessitating a wider scope of words than only emotion words. These words could also likely include those that are searched for when people experience these emotions and can vary from words related to entertainment (e.g., "movies," "Netflix," "music"), social relationships (e.g., "family," "friends"), well-being (e.g., "thankfulness"), to actions taken (e.g., suicide) when people experience negative emotions.

For example, in the Gallup data (Helliwell et al., 2024), the measurement of positive affect (similar to our happiness measure) is defined by the average of three positive affect measures: "laughter", "enjoyment", and "doing interesting things". These measures are obtained from responses to the following three questions: "Did you smile or laugh a lot yesterday?", "Did you experience the following feelings during A LOT OF THE DAY yesterday?" How about Enjoyment?" and "Did you learn or do something interesting yesterday?". While these questions capture aspects of positive affect, the associated keywords may not necessarily reflect how people search for related content online.

Information seeking is a fundamental human drive, arising when people face perplexing situations or become aware of gaps in their knowledge, kindling a desire to fill those voids of understanding (Ford et al., 2018). However, because querying search engines is a goal-oriented solitary activity, there's a fundamental difference in how bridging informational gaps must be examined relative to conventional surveys. To effectively explore positive emotions through search engine queries, survey questions probing affirmative sentiments must be translated into information-seeking terms that capture the nuances of each distinct emotion when submitted to a platform. Indeed, second-person-centric survey questions (e.g., "Did you do or learn something interesting yesterday?") will be translated into instances of their first-person-centric equivalences ("hobbies to explore nearby?"; "movies screening today?"; "concert nearby today?").

For example, using Gallup data, you will not have a Google query such as "Did you smile or laugh yesterday?". Rather, the query could be, "Which movies are showing?" which relates to activities undertaken when experiencing positive emotions. This must be considered when establishing emotion keywords to extract from Google Trends™ to measure experienced happiness.

Information-seeking queries are more likely to include negative emotion words when a person is experiencing an emotion reflecting negative affect, and the search is seeking

information on how the negative affect can be changed to a positive emotion (Ford et al., 2018).

In the Gallup data, negative affect is defined and measured as the average of three negative affect measures. They are "worry", "sadness", and "anger", respectively, the responses to "Did you experience the following feelings during A LOT OF THE DAY yesterday? How about Worry?" "Did you experience the following feelings during A LOT OF THE DAY yesterday? How about Sadness?" and "Did you experience the following feelings during A LOT OF THE DAY yesterday? How about Anger?" (Helliwell et al., 2024).

Therefore, we need to consider "queries seeking information on experienced negative affects" to measure happiness with information-seeking queries to construct a happiness index. If we translate the negative affect (emotions) to an information-seeking question, it could be "What can I do to decrease my anger/worries/sadness? What can I do to minimise the experience of negative emotions? or What can I do to be happy?", thereby maximising positive emotions.

To compile our Google Trends™ dataset, we data-mined the emotion-specific Google queries according to the list of 69 words (see Table 1) derived from the literature and theory (Sects. 2.1 to 2.3) for the UK for the period from January 2011 to December 2023 at a daily frequency. We extracted the data using the Gtrends library in R.

In the initial construction phase, we collapse the daily data to weekly data, summing the observations for the period coinciding with the availability of ONS data per week, from 5 January 2020 to 1 October 2023, giving us 196 observations.

3.1.2 Secondary Datasets – Survey and Big Data

3.1.2.1 Survey Data We considered samples of high-frequency survey data that measure happiness dynamics as a source to validate our new Google Trends™ happiness index. High-frequency survey data measuring happiness is scarce and mainly limited to the US[1] and the UK, although we also have access to the Dutch Time Use survey data. Therefore, the availability of the UK and Dutch data directed our choice of countries in our exploratory analyses.

To address our first research question, we chose the UK's Office for National Statistics (ONS) survey data for the period from 2020 to the end of 2023, with a weekly frequency, which is publicly available. Additionally, as the main language in the UK is English, it allows us to concentrate on only one language.

Specifically, the data we used from ONS forms part of the UK tracking their progress across 10 domains of national well-being, including personal well-being, relationships, and health. Within these 10 domains are 44 indicators of national well-being, including people rating their life satisfaction, happiness, anxiety and whether their lives are worthwhile. We rely on the question where adults aged 16 years and over were asked to rate how happy they felt yesterday on a scale from 0 to 10, where 0 was "not at all" and 10 was "completely" (ONS, 2023). From this point forward, it is referred to as True Happiness.

For the ONS data, we have 196 observations measuring happiness from 5 January 2020 to 1 October 2023, which we use as the outcome variable in training our models to predict happiness. We include the entire period in our analysis to maximise the number of

[1] We did not choose the US since the Gallup American Time Use Survey data is not available unless we incur significant costs.

observations. While we acknowledge the presence of a structural break in the data during the COVID-19 pandemic, we opted to use the whole time series to enhance the likelihood of achieving an accurate model fit. This decision allows us to test our methodology and derive an equation to estimate happiness with high accuracy.

To address our second research question, we use happiness aggregated at the daily level from the Dutch Time Use survey (Bakker et al., 2020) as an outcome variable to explore whether the same basket of emotion words selected for the UK translated into a different language (Dutch) with the same derived weights can also successfully estimate happiness in another country. To test the robustness of our index, we calculated the error metrics between the Dutch Time Use data, which spans the period 2011 to 2021 and our derived happiness measure for the Netherlands. However, there are limitations to the Dutch Time Use survey data, which can lead to a decrease in fit statistics in later years. Since 2020, there have been high levels of missingness per day or very few observations, limiting the representativeness of the data.

3.1.2.2 Big Data We also use Big Data to validate our UK happiness index in the form of the World Health Organisation's (WHO) Early AI-supported Response with Social Listening (EARS) daily dataset. Initially, EARS was used to show real-time information about how people were talking about COVID-19 online. The data was compiled so that the WHO could better manage the situation as the infodemic and pandemic evolved. Although it did not measure happiness, it did measure mental health and loneliness, which we determined from Sect. 2 should have some relationship with happiness. Specifically, we use document 12, representing the number of documents per day for the category mental health and document 33, representing the number of documents per day for the category loneliness.

3.2 Methodology

In this section, we explain the methodology followed to derive our Google Trends™ happiness equation to estimate happiness at a country level.

3.2.1 Correlation to Decrease the Number of Words

After we mined all 69 words from Google Trends™ (Sect. 3.1.1), we tested correlations between weekly measures of the UK's True Happiness and the positive and negative affect emotion words extracted from Google Trends™. We remind the reader that the basket of words includes positive and negative words since positive and negative affect are not the inverse of one another but rather independent and discrete dimensions of affect. At the same time, a person can experience both emotions, which complicates the measure of happiness using information-seeking queries. Suppose the negative affect words are highly correlated with True Happiness (negatively). In that case, we can assume that it has an inverse relationship to True Happiness and is closely related to the measure of positive affect (although the relationship is negative). It can ultimately be used in constructing a happiness index since negative affect drives action, i.e., people are more likely to search for solutions, causes, or validation when experiencing distress (e.g., "how to deal with anxiety"). In addition to capturing positive emotions, we also include search actions which will be undertaken when people experience positive emotions – for example, entertainment-related searches (e.g., movies, Netflix, music), social relationship-related words (e.g., family, friends, friendships) and well-being-related words (e.g., thankfulness).

We selected words statistically significantly correlated to True Happiness, leaving us with 42 words. Since we aim to predict happiness using the most important emotion keywords, we continue the process of decreasing the number of words, training a model using eXtreme Gradient Boosting.

3.2.2 eXtreme Gradient Boosting (XGBoost)

To identify the most important words (features) for predicting happiness, we employed the eXtreme Gradient Boosting (XGBoost) algorithm, a highly efficient and scalable machine learning method that implements gradient boosting for decision trees. XGBoost operates within the gradient boosting framework, where models are developed sequentially, and each new model is trained to correct the errors of its predecessor. This iterative process continues until a robust predictive model is achieved. XGBoost is specifically designed for high performance and speed, utilising optimisation techniques that enable parallel and distributed computing, making it well-suited for large datasets. It also incorporates regularisation methods (L1 and L2) to mitigate the risk of overfitting.

XGBoost has proven to be more accurate than other methods. For example, Abdurrahim et al. (2020) compared various predictive modelling algorithms and found that XGBoost achieved the highest accuracy score when compared to methods like random forest, decision trees, naive Bayes classifier, and logistic regression.

Therefore, our XGBoost model is defined in Eq. (1) as:

$$F_M(x) = F_0 + v\beta_1 T_1(x) + v\beta_2 T_2(x) + \cdots + v\beta_M T_M(x) \tag{1}$$

where M is the number of iterations. The gradient boosting model is a weighted $\left(B_1 \ldots \beta_M\right)$ linear combination of simple models $(T_1 \ldots T_M)$. $F_M(x)$ is the True Happiness, also known as the target variable, measured weekly from 2020 to 2023, with 196 observations. From September 2020, data was only reported every second week. We imputed the data for this period. The independent variables (features) are the 42 words selected through the correlation exercise.

Model evaluation uses metrics to analyse the model's performance, i.e., how well the model generalises future predictions. Machine learning metrics include Accuracy, Precision, Recall and F1 score in classification problems with a discrete, often binary outcome variable. However, we make use of the Mean Absolute Error (MAE), the Mean Squared Error (MSE), and the Root Mean Square Error (RMSE) since our outcome variable is continuous.

3.2.3 Weighting (ElasticNet)

Basic econometric methods are not an option to predict True Happiness as we have many independent variables (features) that are highly correlated (multicollinearity). Working with weekly data of the ONS only available for the period 2020 to 2023 limits our observations to 196. Therefore, using an estimation technique such as OLS to determine which words significantly predict True Happiness was not an option, as the low number of observations means we have insufficient degrees of freedom, and we are challenged by multicollinearity.

Therefore, we turned to ElasticNet, which is a regularised regression ML technique that incorporates both L1 (Lasso) and L2 (Ridge) regularisation penalties into its objective function. It is designed to handle many features in smaller datasets. By combining the

L1 and L2 penalties, ElasticNet can achieve both feature selection and parameter shrinkage, making it particularly useful in scenarios with highly correlated predictors (highly correlated words, as in the present study). In addition, it mitigates the risk of overfitting by shrinking less important coefficients and potentially setting some to zero (like Lasso); however, it does not fully eliminate the risk, particularly when working with small datasets as in the current study.

The success of ElasticNet depends heavily on the choice of its hyperparameters, namely:

Alpha: The mixing parameter between Lasso (L1) and Ridge (L2) penalties. Alpha = 1 corresponds to Lasso, while Alpha = 0 corresponds to Ridge. In our analysis, we used the default Alpha parameter of 0.5.

Lambda (or L1_ratio): The regularisation strength. Higher values mean more regularisation.

Due to the small sample size, K-fold cross-validation is crucial to ensure the model generalises well to new data. This technique involves partitioning the data into K subsets and then iteratively training the model on K-1 subsets while using the remaining subset for testing.

In addition to using K-fold cross-validation to mitigate the potential of overfitting, we also introduced other measures, as seen in Sect. 3.2.4. Nonetheless, we are aware of the potential of reduced generalisability of the models when applied to unseen data when interpreting results.

We use the estimated coefficients of the features from ElasticNet as weights to derive our happiness index, which estimates country-level happiness.

3.2.4 Steps Taken to Guard Against Overfitting

Due to the risk of overfitting when using relatively smaller datasets, we included the following steps in both the XGBoost and the ElasticNet models. First, in terms of the XGBoost, we used i) a strict feature selection, limiting the number of features to only those words statistically significant to the survey happiness measure, i.e., "True Happiness"; ii) randomised sample selection throughout; iii) small tree depth of 3 to limit the complexity, which normally improves generalisation; iv) specified an early stop criteria by monitoring the RMSE; if there were no improvements after five iterations, the training process was stopped. In terms of ElasticNet, as mentioned above, it is adapted explicitly to small sample sizes to prevent overfitting with its regularisation (Lambda and Alpha) and adds penalties to the coefficients (shrinking the coefficients). Furthermore, the features included in the ElasticNet models underwent a two-stage selection process whereby only statistically significant words were included in the XGBoost – and using the XGBoost, the features were further reduced to only include the most important features. The strictly selected features limit the risk of including noise in the estimation. We used k-fold cross-validation to assess model performance and ensure it generalises well.

For both the XGBoost and ElasticNet, we monitored the performance metrics (RMSE) of the training and test sets. The metrics were very similar, indicating a good fit – if there were large gaps, it might have indicated overfitting.

3.2.5 Robustness Checks

Lastly, we rely on unseen data to test the robustness of our derived Google Trends™ happiness index. To test the time invariance of our derived index, we applied our happiness

equation to quarterly ONS data to determine if the trends captured using weekly data and quarterly are similar.

Additionally, we test the robustness of our Google Trends™ happiness index using Big Data. Here, we relied on daily data from EARS: document 12, representing the number of documents per day for the mental health category and document 33, representing the number of documents per day for the category loneliness, to test the correlations between our daily derived happiness index and these measures.

4 Results and Analysis

4.1 Constructing a Happiness Index Using Google Trends™ for the UK

4.1.1 XGBoost Initial Model on Search Terms' Relative Importance

To determine the most important words from the 42 words identified in Sect. 3.2.1, we use XGBoost (Sect. 3.2.2). The outcome variable (Label) is the UK's True Happiness (happiness as reported in the ONS survey data). We use the regression option within the XGBoost algorithm since True Happiness is a continuous variable.

We start by randomly splitting the data into a training and testing dataset with an 80:20 split on all data, with the evaluation done on the unseen testing data. We used the random split method to ensure that both datasets represent the overall distribution.

To train the model, we initially used all the default settings of the parameters of the XGBoost algorithm. Next, we predict True Happiness, evaluate the model according to the fit metrics, and refine the parameters to optimise the model's performance. We found the optimal tree depth was three.[2] We set the number of iterations to 100, with a termination clause added to stop the algorithm if the RMSE does not decrease after 5 iterations.

The RMSE evaluating the fit reached its lowest value of 0.11268 at the 32nd boosting cycle, after which the training stopped due to the "early stop" criteria of 5 if the RMSE did not improve.

The evaluation metrics for our XGBoost model show that all measures of fit reveal small errors, indicating a good-fitting model. For the XGBoost, the MSE is 0.013, the MAE is 0.083, and the RMSE is 0.113.

Following the results from the XGBoost model, those words with gains of more than 0.01 were retained, leaving us with 26 words.

From the 26 words, we found that "sad" was the most important word with a gain of 0.2571, followed by "headache" (gain of 0.1657), "depressed" (gain of 0.0547) and "music" (gain of 0.0423). It is interesting to note that negative emotion words indeed are significant predictors of True Happiness. This agrees with our earlier discussion that using information-seeking Google queries will most likely lead to finding the negative queries important. Positive words that are important predictors include "well-being" (gains = 0.0401), "love" (gains = 0.0366) and "great" (gains = 0.0236).

[2] In XGBoost, tree depth refers to the maximum depth of individual trees in the ensemble. A depth of three means that each tree in the boosting process was limited to three levels of splits, optimising the trade-off between model complexity and generalisation while mitigating overfitting.

4.1.2 ElasticNet Weighting and Aggregation

As mentioned in Sect. 3.2.3, we used an ElasticNet linear regression algorithm, a machine learning approach, to estimate the coefficients. Once the coefficients are determined, we use these as weights in the equation to estimate the happiness levels in countries.

To train the ElasticNet model, we randomly split the data into a training and testing dataset with an 80:20 split on all data. We initially started by using all the default parameters for ElasticNet with an Alpha and Lambda of 0.5. After conducting the fivefold cross-validation process, we identified that the optimal (best) Lambda was 0.000193946, which minimised prediction error and indicated a moderate level of regularisation, which is in line with the complexity of the dataset.

The evaluation metrics indicate a good fit with an MAE of 0.0690, an MSE of 0.0117, and an RMSE of 0.0938, indicating that our model predicts True Happiness well. The scatter plot in Fig. 1 confirms the good fit of the *Predicted Happiness* versus True Happiness scores.

Therefore, the generic equation to estimate happiness in the UK is as follows:

$$GNH\_GT = \beta_0 + \alpha_1 * \beta_1 + \cdots \cdots \cdots + \alpha_{26} * \beta_{26} \tag{2}$$

where $\alpha_{11} \cdots \alpha_{26}$ represents the 26 words as determined by our XGBoost model to be the most important words in predicting the outcome variable, True Happiness, and $\beta_1 \cdots \beta_{26}$ represent the weights as determined by the coefficient results from the ElasticNet linear regression model.

For example, in the newly derived equation to estimate happiness, the weights for "sad", "headache", and "depressed" are −0.1324, −0.1230 and −0.0228, respectively.

Applying the newly derived equation, we estimate happiness in the UK. Figure 2 shows the *Estimated Happiness* versus True Happiness (ONS weekly survey data). Evaluating the fit, the RMSE is 0.0940, which indicates a very good fit.

4.1.3 Results from Robustness Checks and Validation Exercise

As mentioned in Sect. 3.2.4, we test the time invariance of our Google Trends™ happiness index by applying our derived Eq. (2) to quarterly ONS data. We observe that the trends

Fig. 1 Predicted happiness vs true happiness for the UK

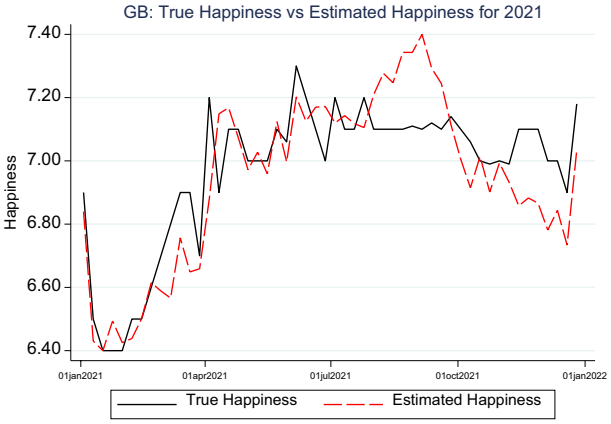

Fig. 2 Estimated happiness vs true happiness from the UK ONS data

captured using weekly data are also reflected in the quarterly data. Table 2 shows the correlation between the estimated happiness indices using weekly and quarterly data, which is strong and significant at 0.7000 (p = 0.000).

Additionally, we validate our Google Trends™ happiness index using Big Data. Here, we tested the correlation between our derived happiness index and daily data from EARS: document 12, representing the number of documents per day for the mental health category and document 33, representing the number of documents per day for the category loneliness. Table 2 shows statistically significant and negative correlations of -0.31 (mental health) and -0.25 (loneliness). The negative correlations are as expected.

Considering the correlation results in Table 2, we are confident that our equation yields consistent results regardless of the data frequency (e.g., weekly or quarterly), and it is robust when validated against other well-being measures.

4.2 Constructing a Happiness Index Using Google Trends™ for the Netherlands

4.2.1 Estimating Happiness in the Netherlands Using the UK-Derived Equation

This section reports the results of our second research objective. Here, we explore whether the same basket of words used in the UK index translated into a different language (Dutch) with the same derived weights can also successfully estimate happiness in another country. As mentioned in Sect. 3.1.2, we translated the basket of 26 words determined by our XGBoost model into Dutch and applied our equation to the extracted Dutch words. To test the validity of our *Estimated Happiness*, we correlate it with True Happiness as recorded in the Dutch Time Use survey data (Bakker et al., 2020).

Table 2 Correlation of Google Trends™ happiness index using quarterly ONS data and high-frequency data from EARS. *Source*: Authors own calculations

Measure	Estimated happiness
Estimated happiness	1
ONS quarterly happiness	0.7000 ***
EARS–doc 12 (mental health)	−0.3121 ***
EARS–doc 33 (Loneliness)	−0.2425***

Table 3 shows that the estimated happiness using the Dutch equivalent of our English words and weights (equation to estimate happiness for the UK) is statistically significantly correlated to happiness recorded in the Dutch Time Use Survey. In 2011, the correlation was strong, at 0.5742, though it performed much weaker in 2020, with the correlation being 0.2589.

Figures 3 and 4 show the *Estimated Happiness* against True Happiness measured by the Dutch Time Use survey data. The fit statistics show an RMSE of 0.08 for 2011; however, the fit weakens considerably with an RMSE of 0.43 for 2020. The results are similar to those revealed using correlation analysis. The result of a weaker fit in 2020 is surprising as we expected the estimated happiness to show a better fit for 2020, as the period coincides with that used to derive the UK equation. However, a likely

Table 3 Correlation of Google Trends™ happiness index in Dutch correlated to Dutch Time Use survey data. *Source*: Authors own calculations

Measure	Estimated Happiness
Estimated happiness	1
Dutch time use survey happiness—2011	0.5742***
Dutch time use survey happiness—2020	0.2589***

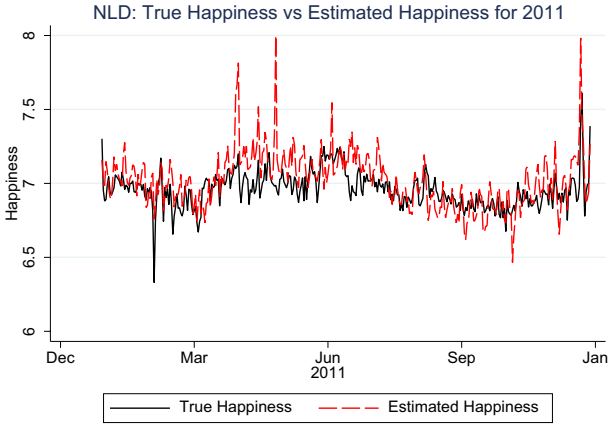

Fig. 3 Estimated happiness vs true happiness from the netherlands dutch time use survey data (2011)

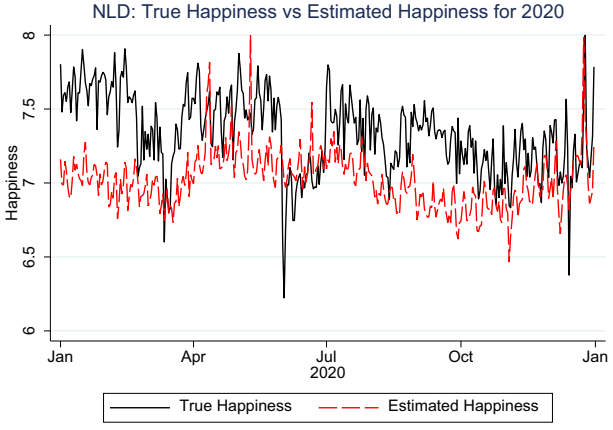

Fig. 4 Estimated happiness vs true happiness from the netherlands dutch time use survey data (2020)

explanation is Dutch Time Use data quality, as it has deteriorated over time, with significantly fewer respondents and many missing observations since 2020.

Considering our results, we re-estimate our happiness equation for the Netherlands in the next section by incorporating country-specific words. To predict True Happiness using our machine learning algorithms in the Netherlands, we use the Dutch Time Use Survey data for the period 2011 and 2012 (during these two years, the quality of the Time Use data was good).

4.2.2 Constructing a happiness index using Google Trends™ for the Netherlands, including country-specific emotion words

This section reports the results of our third research objective. Here, we use the same set of words as those we used for the UK but add relevant country-specific words such as "perfect" and "fijne" to reflect emotion words used in Dutch. We started our analyses with the same initial 69 emotion words (refer to Table 1) translated into Dutch, adding the country-specific words. Subsequently, we used correlation analysis to establish which of the extracted words were significantly correlated to the Dutch Time Use survey's happiness measure (Bakker et al., 2020) and reduced the initial basket of words to 47 words.

To determine the most important words from the 47 words identified above, we use XGBoost (Sect. 3.2.2). The outcome variable (Label) is the Netherlands' True Happiness (Dutch Time Use survey's happiness measure). Using XGBoost, we follow the method explained in Sect. 4.1.1 by randomly splitting the data into a training and testing dataset with an 80:20 split on all data, with the evaluation done on the unseen testing data.

The evaluation metric for our XGBoost model reveals a small error, indicating a good-fitting model. For the XGBoost, the RMSE is 0.0501. Following the results from the XGBoost model, those words with gains of more than 0.01 were retained, leaving us with 23 words. The most important words included, among others, "fijne" (gain = 0.1837), "kanker" (gain = 0.1304), "hoofdpijn" (gain = 0.0785), and "dood" (gain = 0.0638).

We use the most important words as features to train the ElasticNet regression model to predict True Happiness in the Netherlands. The evaluation metrics demonstrate strong predictive performance with an MAE of 0.0345, an MSE of 0.0025, and an RMSE of 0.0504, indicating that our model predicts True Happiness well.

We use the estimated coefficients to weight the features (words) to derive the happiness equation for the Netherlands. Figure 5 shows the True Happiness (from the Dutch time use survey data) and the *Estimated Happiness* for the year 2011.

A visual inspection suggests a good fit. To evaluate the accuracy of our *Estimated Happiness* versus True Happiness, we calculated the RMSE. The RMSE for True Happiness versus *Estimated Happiness* is 0.0504, indicating a smaller error compared to the initial derived happiness equation, which had an RMSE of 0.0823, where no country-specific words were included (see Sect. 4.2.1). Furthermore, the fit is markedly better than the one we attained for 2020, with an RMSE = 0.43.

Therefore, we can conclude that adding country-specific words decreases errors and improves the accuracy of happiness estimations using information-seeking query data extracted continuously from Google Trends™.

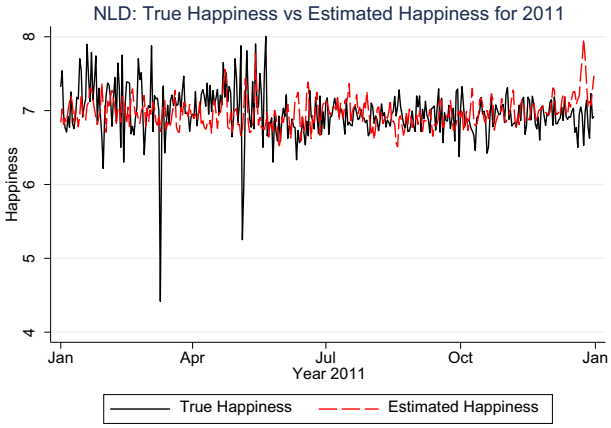

Fig. 5 Estimated happiness vs true happiness from the netherlands dutch time use survey data with country-specific emotion words

5 Conclusions

In this paper, we constructed and validated a real-time happiness index using search query data based on emotion keywords, which we extracted from Google Trends™, representing the first of its kind to our knowledge. Our Google Trends™ happiness index for the UK combines 26 words, each with its own weight as determined by our ElasticNet linear regression machine learning model.

We initially started with carefully curated words suggested to capture positive and negative affect. We extracted the 69 words using Google Trends™ from the UK and correlated those to the weekly UK True Happiness measure obtained from the ONS data. Words significantly correlated to the happiness score were selected for further analysis. We were left with 42 words, which subsequently decreased to 26, given our results from the XGBoost model that determined the most important words (features) in predicting True Happiness obtained from the UK's ONS data (outcome variable). Subsequently, we used the ElasticNet linear regression machine learning model to estimate the coefficients of each of the 26 words in predicting happiness. These coefficients were then applied as weights in our equation to estimate happiness.

To test the time invariance of our Google Trends™ happiness index for the UK, we applied our derived equation to quarterly ONS data. Additionally, we validated our Google Trends™ happiness index using Big Data in the form of EARS: document 12 represents the number of documents per day for the category mental health, and document 33 represents the number of documents per day for the category loneliness. Considering the correlation results, we are confident that our equation yields consistent results regardless of the data frequency (e.g., weekly or quarterly), and it is robust when validated against other well-being measures.

We used data from the Netherlands to explore whether the same basket of words translated into a different language (Dutch) with the same derived weights can also successfully estimate happiness. We translated our 26 English words into Dutch and applied our derived happiness equation. Then, we correlated it with the True Happiness measure from the Dutch Time Use survey and found a statistically significantly strong relationship in 2011 and a weaker relationship in 2020. We also plotted our *Estimated Happiness* versus True Happiness in 2011 and 2020 and calculated the respective RMSEs. The results showed a good fit for 2011 (RMSE = 0.08) and a weaker fit for 2020 (RMSE = 0.43). A plausible

reason is the quality of the Time Use data, which weakened over time with fewer respondents and fewer observations.

Lastly, we followed the same methodology we used for the UK to derive a happiness equation. Here, we used the 69 identified emotion words and added country-specific words. We determined the most important words using XGBoost. We used those words in an ElasticNet algorithm to estimate the coefficient of each word and subsequently applied them as weights in our happiness equation for the Netherlands. Using the derived equation, we estimated Dutch happiness. To test the accuracy of our estimated happiness, we compared it to True Happiness. We calculated the RMSE, which is 0.05, indicating that although emotion words show an overlap between countries, including country-specified words can improve happiness estimations.

Therefore, we successfully showed that information-seeking queries extracted using Google Trends™ can be used to estimate happiness and construct a near real-time happiness index.

The result of our study provides several practical initiatives. First, governments and policymakers can leverage the real-time insights provided by the Google Trends™ happiness index to monitor national mood fluctuations and respond accordingly to mitigate potentially violent and destructive outcomes such as violent riots. Second, since happiness is linked to productivity and economic performance, monitoring happiness trends in real-time can inform labour policies. For example, significant dips in happiness might indicate rising workplace dissatisfaction, burnout, or economic hardship, prompting governments to adjust workplace well-being initiatives or financial support programmes. Third, real-time tracking of happiness can help policymakers anticipate public reactions to major policy decisions. By analysing how past legislative changes, such as the COVID-19 mandates, impacted national happiness, decision-makers can better predict potential resistance to new policies and implement communication strategies to mitigate backlash. Lastly, our study has policy implications for international organisations such as the OECD and the United Nations that aim to measure global well-being. Policymakers should consider local linguistic and cultural variations when designing happiness-tracking frameworks.

It is important to acknowledge two key considerations related to our happiness measure. First, our Google Trends™ happiness index is time-sensitive, requiring intermittent review and confirmation of the selected emotion words to ensure that our derived happiness equation is still accurate in estimating a country's happiness. Moreover, while the composition of happiness equations may vary slightly across countries—each incorporating a few different emotion words—we maintain that these differences do not compromise the validity of the measure. Despite cultural variations in the expression of happiness, the concept itself is broadly recognised across societies. As such, while some individual terms may differ, all happiness equations ultimately converge on the same fundamental outcome: a meaningful representation of well-being.

Second, while we demonstrated the effectiveness of using Google Trends™ to measure real-time happiness in the UK and the Netherlands (English and Dutch), the generalisability of these findings to other significantly different linguistic and cultural contexts remains an open question. Languages differ in how they encode and express emotions, making direct translation of happiness-related terms difficult for capturing cultural nuances. Additionally, internet search behaviours vary across regions due to disparities in digital access, privacy concerns, and the use of alternative search engines. To enhance cross-cultural validity, future research should incorporate expert linguistic reviews to refine translated emotion terms. Computational techniques, such as multilingual word embeddings, could further improve scalability by identifying semantically equivalent emotion words across

languages. A crowdsourced approach involving native speakers could also help refine emotion lexicons, ensuring broader applicability and consistency across different cultural contexts.

Our future research will endeavour to use our Google Trends™ methodology to construct and estimate indices of subjective well-being at a country level in real-time, including life satisfaction. Apart from extending our research agenda, we are committed to the continuous refining and adaptation of our methodology to ensure its robustness across diverse linguistic, cultural, and technological landscapes. We will continue to explore strategies to enhance the flexibility of our approach, including integrating dynamic keyword validation and employing human-in-the-loop validation processes to refine emotion lexicons. Additionally, we will expand our dataset to include a broader range of countries, which will allow us to understand better the influence of digital behaviour on our happiness measure. By actively iterating on our methods, we aim to strengthen the scalability, reliability, and universality of Google Trends™ as a tool for real-time well-being assessment.

Acknowledgements We thank Dr Frederic Boy from Swansea University for his valuable feedback on an earlier draft.

Authors' contributions All authors contributed equally to the study's conception and design.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. We thank the Well-being in a Sustainable Economy Revisited (WISER) project for the funding received through Horizon Europe (Grant agreement ID: 101094546).

Data availability The data used are from Google Trends ™, an open data service offered by Google Inc. (https://trends.google.co.uk/).

Declarations

Conflict of interest Stephanie Rossouw is the economics editor, and Talita Greyling is an associate editor for the Journal of Happiness Studies. The authors declare they have no financial interests.

Ethics approval This article does not contain any studies with human participants or animals performed by any authors.

Consent for publication The authors claim that none of the material in the paper has been published or is under consideration for publication elsewhere.

References

Abdurrahim, Y., Ali, A. D., Sena, K., & Huseyin, U. (2020). Comparison of deep learning and traditional machine learning techniques for classification of pap smear images. *arXiv*, 2009.06366v1. Available from https://arxiv.org/pdf/2009.06366.pdf

Algan, Y., Murtin, F., Beasley, E., Higa, K., & Senik, C. (2019). Well-being through the lens of the internet. *PLoS ONE, 14*(1), e0209562.

Bakker, A., Burger, M., van Haren, P., Oerlemans, W., & Veenhoven, R. (2020). Raise of happiness following raised awareness of how happy one feels: A follow-up of repeated users of the happiness indicator website. *International Journal of Applied Positive Psychology, 5*, 153–187.

Banerjee, S. (2018). How does the world google the internet, anxiety, and happiness? *Cyberpsychology, Behavior, and Social Networking, 21*(9), 569–574.

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). The development and psychometric properties of LIWC-22. Austin, TX: University of Texas at Austin. https://www.liwc.app

Brodeur, A., Clark, A. E., Fleche, S. & Powdthavee, N. (2020). Assessing the impact of the coronavirus lockdown on unhappiness, loneliness, and boredom using Google Trends. arXiv: 2004.12129. Available at https://ui.adsabs.harvard.edu/abs/2020arXiv200412129B/abstract

Bryson, A., Clark, A. E., Freeman, R. B., & Green, C. (2016). Share capitalism and worker well-being. *Labour Economics, 42*, 151–158.

Callegaro, M., & Yang, Y. (2018). The role of surveys in the era of "big data." In D. L. Vannette & J. A. Krosnick (Eds.), *The palgrave handbook of survey research* (pp. 175–192). Springer International Publishing. https://doi.org/10.1007/978-3-319-54395-6_23

Carammia, M., Iacus, S. M., & Wilkin, T. (2022). Forecasting asylum-related migration flows with machine learning and data at scale. *Scientific Reports, 12*(1), 1–16.

Carpi, T., Hino, A., Iacus, S. M., & Porro, G. (2022). The impact of COVID-19 on subjective well-being: Evidence from twitter data. *Journal of Data Science, 21*(4), 761–780.

Crawford, J. R., & Henry, J. D. (2004). The positive and negative affect schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 3*(Pt 3), 245–265.

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D.-W., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research, 97*, 143–156.

Dodds, P. S., & Danforth, C. M. (2010). Measuring the happiness of large-scale written expression: songs, blogs, and presidents. *Journal of Happiness Studies, 11*, 441–456.

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PLoS ONE, 6*(12), e26752.

Engelen, U., De Peuter, S., Victoir, A., Van Diest, I., & Van Den Bergh, O. (2006). Verdere validering van de positive and negative affect schedule (PANAS) en vergelijking van twee Nederlandstalige versies [further validation of the positive and negative affect schedule (PANAS) and comparison of two Dutch versions]. *Gedrag & Gezondheid: Tijdschrift Voor Psychologie En Gezondheid, 34*(2), 89–102.

Foa, R. S., Fabian, M., & Gilbert, S. (2022). Subjective well-being during the 2020–21 global coronavirus pandemic: Evidence from high frequency time series data. *PLoS ONE, 17*(2), e0263570.

Ford, M. T., Jebb, A. T., Tay, L., & Diener, E. (2018). Internet searches for affect-related terms: An indicator of subjective well-being and predictor of health outcomes across US states and metro areas. *Applied Psychology: Health and Well-Being, 10*(1), 3–29.

Greyling, T., Rossouw, S., & AFSTEREO. (2019). *Gross National Happiness.today Index*. Available from http://gnh.today

Greyling, T., & Rossouw, S. (2022). Positive attitudes towards COVID-19 vaccines: A cross-country analysis. *PLoS ONE, 17*(3), 0264994.

Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., & Wang, S. (Eds.). (2024). *World Happiness Report 2024*. Appendix 1: Statistical Appendix for Chapter 2. University of Oxford: Wellbeing Research Centre

Houghton, S., Boy, F., Bradley, A., James, R., Wardle, H., & Dymond, S. (2023). Tracking online searches for gambling activities and operators in the United Kingdom during the COVID-19 pandemic: A google trends™ analysis. *Journal of Behavioral Addictions, 12*(4), 983–991.

Iacus, S. M., Porro, G., Salini, S., & Siletti, E. (2015). Social networks, happiness and health: From sentiment analysis to a multidimensional indicator of subjective well-being. *ArXiv*. https://doi.org/10.48550/arXiv.1512.01569

Iacus, S. M., Porro, G., Salini, S., & Siletti, E. (2022). An Italian composite subjective well-being index: The voice of twitter users from 2012 to 2017. *Social Indicators Research, 161*, 471–489.

Jovanović, V., Joshanloo, M., Martín-Carbonell, M., Caudek, C., Espejo, B., Checa, I., Krasko, J., Kyriazos, T., Piotrowski, J., Rice, S. P. M., Junça Silva, A., Singh, K., Sumi, K., Tong, K. K., Yıldırım, M., & Żemojtel-Piotrowska, M. (2022). Measurement invariance of the scale of positive and negative experience across 13 countries. *Assessment, 29*(7), 1507–1521.

Kahn, J. H., Tobin, R. M., Massey, A. E., & Anderson, J. A. (2007). Measuring emotional expression with the linguistic inquiry and word count source. *The American Journal of Psychology, 120*(2), 263–286.

Kasser, T., & Ryan, R. M. (1996). Further examining the American dream: Differential correlates of intrinsic and extrinsic goals. *Personality and Social Psychology Bulletin, 22*, 280–287.

Kercher, Kyle. (1992). Assessing subjective well-being in the old-old: The PANAS as a measure of orthogonal dimensions of positive and negative affect. *Research on Aging, 14*(2), 131–168. https://doi.org/10.1177/0164027592142001

Kim, E. S., Kubzansky, L. D., & Smith, J. (2015). Life satisfaction and use of preventive health care services. *Health Psychology, 34*(7), 779–782.

Mackinnon, A., Jorm, A. F., Christensen, H., Korten, A. E., Jacomb, P. A., & Rodgers, B. (1999). A short form of the positive and negative affect schedule: Evaluation of factorial validity and invariance across demographic variables in a community sample. *Personality and Individual Differences, 27*, 405–416.

Murtin, F., & Salomon-Ermel, M. (2024). Nowcasting subjective well-being with google trends: A meta-learning approach. OECD papers on well-being and inequalities. Working Paper No.27. Available from https://doi.org/10.1787/4ca48f7c-en

ONS (2023). Measures of national well-being: quality of life in the UK – May 2023. Available from https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/datasets/measuringnationalwellbeingdomainsandmeasures

Piekalkiewicz, M. (2017). Why do economists study happiness? *Labour, 28*(3), 361–377.

Rossouw, S., & Greyling, T. (2020). Big Data and Happiness. Invited chapter for the Handbook of Labor, Human Resources and Population Economics. Edited by Klaus F. Zimmermann. https://doi.org/10.1007/978-3-319-57365-6_183-1

Rossouw, S., & Greyling, T. (2024). Big data and happiness. In H. Brockmann & R. Fernandez-Urbano (Eds.), *Encyclopedia of happiness, quality of life and subjective wellbeing* (pp. 310–315). Edward Elgar Publishing. https://doi.org/10.4337/9781800889675.00051

Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and Negative affect schedule (PANAS). *Journal of Cross-Cultural Psychology, 38*(2), 227–242.

Watson, D., & Clark, L. A. (1994). The PANAS-X: manual for the positive and negative affect schedule - expanded form. Available from https://iro.uiowa.edu/esploro/outputs/other/The-PANAS-X-Manual-for-the-Positive/9983557488402771

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063–1070.

Williams, S., & Shiaw, W. T. (1999). Mood and organisational citizenship behavior: The effects of positive affect on employee organisational citizenship behavior intentions. *Journal of Psychology, 133*, 656–668.

Zevon, M. A., & Tellegen, A. (1982). The structure of mood change: An idiographic nomothetic analysis. *Journal of Personality and Social Psychology, 43*(1), 111–112.