Wells Fargo Campus Analytics Challenge Official Rules

PLEASE READ CAREFULLY—THESE OFFICIAL RULES ADDRESS YOUR LEGAL RIGHTS

THIS IS A JUDGED CHALLENGE. NO ENTRY FEE OR PURCHASE IS REQUIRED TO PARTICIPATE. YOU MUST ENTER THIS CONTEST WHILE IN THE UNITED STATES.

1. **Overview:** At Wells Fargo, our data scientists play a key role in driving innovative and meaningful insights that enable our lines of business to provide a world-class experience to our stakeholders. Data scientists who are developing models to classify and/or extract information from text data have enabled the bank to identify key insights and quantify customer feedback, a process that used to be extremely manual. The Campus Analytics Challenge 2019 ("Challenge") puts you in the role of a natural language processing (NLP) data scientist and calls you to develop a topic model beyond a common approach (latent Dirichlet allocation, LDA), as described below. The dataset is small enough that you should be able to work with it on a standard laptop.

The data you will be working with are descriptions of open datasets published by NASA. These data also exhibit a high degree of overlap between high-level topics.

To help get your creative juices flowing, we encourage you to explore topic modeling research literature and beyond, as you may find a creative approach in other sub-fields of NLP such as summarization or sentence simplification.

2. **Challenge Schedule:** Challenge runs according to the schedule below on https://www.mindsumo.com/companies/wells-fargo (the "Challenge Website").

Challenge Starts	2:00 p.m. Eastern Time ("ET") on 10/28/19
Challenge Ends	2:00 p.m. ET on 12/2/19
Submissions Judged	12/3/19 – 12/16/19
Potential Winners Notified (winning subject to verification)	12/17/19 (on or about)

3. **Eligibility:** This Challenge is sponsored by Wells Fargo Bank, N.A. ("Sponsor" or "Wells Fargo") for full-time or part-time students, 18 years of age or older at the time of entry, who are enrolled in any higher education degree program on campus or online at colleges or universities in the United States, including students attending two- and four-year programs, technical and vocational schools, junior and community colleges, as well as graduate and professional education students (collectively "Students"). Employees of Wells Fargo or MindSumo, Inc. and their respective parents, divisions, affiliates, subsidiaries, their promotional or marketing agencies, government entities and public officials, and their immediate family members (parent, child, sibling and spouse) and persons living in the same households of each such employee (whether related or not) are not eligible. **To be eligible to receive any prize, potential winners must have a valid U.S. tax identification number and meet all the eligibility requirements at the time the prize is awarded.** Potential winners may be required to provide Sponsor with proof that they meet the eligibility requirements for this Challenge. Void where prohibited by law.

4. **The Challenge:** As noted above, the Challenge is for you to develop a topic model beyond a common approach (latent Dirichlet allocation, LDA). Your topic model ("Solution") must meet: (a) the Challenge Objectives, (b) follow the Challenge Instructions and Requirements, and (c) incorporate the Key Deliverables, each of which is described in detail below.

5. **Challenge Objectives:** Topic models, methods that extract themes from unstructured text data, often provide a first layer of insights. Two main requirements regarding model output present a challenge to traditional topic models:

    a) Correlations between topics present in many of the corpora violate the assumption of topic independence made by many topic models, including the popular latent Dirichlet allocation ("LDA").

    b) Hierarchical models, which provide insights on multiple levels along the spectrum of broad to detailed, are more useful than a single, coarse, high-level segmentation of documents into large thematic bins.

Therefore, the two main objectives in this Challenge are for your Solution to appropriately handle correlated topics and to generate subtopics (in addition to topics).

6. **Challenge Instructions and Requirements:** When creating your Solution, you may use **a novel combination** of existing machine learning and/or statistical methods, **or develop your own** novel method in order to extract and/or represent thematic information from the text. Either way, the output needs to include:

    a) a distribution of topics over documents, and
    b) a distribution of words over topics.

Additionally, your Solution must meet the following requirements:

    a) You must use Python 3.

    b) You must provide citations and sources.

    c) Your method must be dynamic, i.e. easy to re-run repeatedly as new dataset descriptions come in, as opposed to being fixed on the set of dataset descriptions provided. An implication of this requirement is that the topics identified in the first run should be stable enough that new descriptions can either cleanly merge with existing topics, or form a new topic of their own.

    d) You must develop a hierarchical model that provides both topics and sub-topics present in the dataset. Your model should have at least two levels (topics and sub-topics), but may have more.

    e) You must produce an LDA model as a baseline against which to compare your method. The comparison may be qualitative, although bonus points will be given for more rigorous quantitative comparisons.

f) Your topic model should not use nor rely on any of the structured fields in the dataset. You may only use the structured fields to evaluate your Solution, **after** the topic model produces its results.

7. <u>Challenge Suggestions</u>:

a) You may use LDA as a component of your novel method, although be warned that the dataset has been modified such that LDA alone will not produce good results. Instead, we encourage you to ***think broadly and creatively.***

b) You may use any clustering, dimensionality reduction, or other algorithm families. Please note that among other criteria, you will be evaluated on whether your selections of methods are appropriate for unstructured data, i.e. that using unstructured data as input does not violate any assumptions.

c) You may stratify the data by date issued or modified, or choose not to stratify. You may not use the date stratification alone as the top layer of your topic identities.

8. <u>Key Deliverables</u>:

<u>Deliverable 1</u>: *Your results*

a) A table of your results assigning topics to a list of dataset description identifiers in the format shown here. If your method is probabilistic, you should add a third column that contains a list of the probabilities associated with each review. More generally, this column may contain a measure of your confidence for the assignment of the dataset to the topic.

Very important note: The dataset_id should be numeric, with a range from 1 to 24395, in the **same** order as the original .json file. If this is not the case, the submission will be automatically disqualified.

Please note that if a dataset description is in more than one topic, that dataset description will have multiple rows. Therefore, the number of rows in this table will likely be larger than the number of dataset descriptions.

topic	sub_topic	dataset_id	some_confidence_measure

b) Similarly, please include in your submission a table of your results assigning words (and/or other features such as n-grams) to the lists of topics in which they appear. For this table, the number of rows will be equal to the number of features.

The confidence measure is optional but highly encouraged. It could be confidence intervals, probabilities, feature weights, or any other measure that makes sense in the context of your method.

feature	topics	sub_topics	some_confidence_measure

Deliverable 2: *Your method*
A description of your approach delivered as:
  a) A visual description (flow chart or similar) of the path of the data through your pipeline. Note the areas where your approach is novel.
  b) A few paragraphs describing the rationale behind your method. In other words, how does your method address the challenges of topic overlap and high-resolution insights? Make sure to acknowledge and justify the assumptions of the methods you choose to use (maximum 800 words).

Deliverable 3: *Your code*
  a) Well-commented code that is operational and can be run using the data provided and generating the output of your approach.
  b) An environment configuration file that lists the names and versions of the libraries you used.

Deliverable 4: *Your end-user communication tools*
  a) An end-user visualization or tool that highlights your insights about themes in the corpus.
  b) A figure legend for your end-user visualization or tool. Make sure to include a short description of both strengths and limitations of your method (maximum 200 words).

  Tips for deliverable 4:
  i. Display this visualization side-by-side with a visualization of your baseline traditional LDA model, allowing for a qualitative comparison.
  ii. Your visualization must highlight insights produced by your method, but were undetected by LDA.
  iii. Your visualization may be interactive.
  iv. We encourage you to go beyond using standard tools such as pyLDAvis, which is better suited as a development-stage visual aid, rather than as a tool to communicate insights.
  v. Be mindful of over-generalizing or misleading with your visualization, i.e. please explicitly address areas of potential uncertainty inherent in your method that may limit the conclusions you can make from the visualization.

 9. **Entering/Submitting Your Solution:** To submit your Solution and enter the Challenge before the Challenge Submission Deadline (see Rule 2), you must do the following:
  a) Register for a free MindSumo.com account. If you are already a MindSumo registrant, you can proceed to step **b)**. To register with MindSumo, you will be required to agree to the MindSumo Terms of Use available at www.mindsumo.com/terms. The MindSumo Terms of Use apply to this Challenge. Capitalized terms used in these Official Rules will have the meaning given to them in the MindSumo Terms of Use, unless otherwise noted.
  b) Find the Wells Fargo Challenge in the Challenge Listings or you can go directly to the Challenge Website at https://www.mindsumo.com/companies/wells-fargo. (If you are a new registrant, you will be directed automatically to the Challenge Listings upon signing up.)

c) Submit your Solution by typing it directly into the Challenge Website or uploading it as a separate file. You may edit your Solution on the Challenge Website until the Challenge Submission Deadline, or you can compile your Solution and upload it as individual files or as a zip file. For accessibility, PDFs can be converted from text to speech (https://www.adobe.com/enterprise/accessibility/pdfs/acro6_cg_ue.pdf). The Challenge will also be available to Students via audio file (mp3).

By entering, you warrant and represent the following with respect to your Solution: (a) you own or control all rights in and to the submission; (b) the Solution has not been and will not be assigned or transferred to another party; (c) the Solution is yours to give to Sponsor per these Official Rules; (d) the Solution is your own creation; and (e) nothing in the Solution will infringe on any rights of any third parties.

10. **Challenge Judging:** Your submission will be judged based on your responses to each Key Deliverable of the Solution based on the criteria in Rule 11 below. Each judge may award a **maximum of 170 points** for any given Solution. Each submitted Solution will be initially screened for eligibility and appropriateness. At Sponsor's sole discretion, Solutions which do not specifically address one or more Key Deliverables, are incomplete, do not pass the screening, and/or are deemed to be negative, offensive or controversial in any way will be disqualified from the Challenge. Sponsor makes the final determination as to which submissions are eligible to take part in the Challenge.

Round 1 Judging: All eligible submissions will be judged by a panel of judges from MindSumo using the judging criteria below. The Solutions with the top 10 scores will be selected as potential finalists ("Finalists").

Round 2 Judging: Each of the Finalist's Solutions will then be judged by a second panel of judges from Wells Fargo and MindSumo based on the judging criteria below to select up to five potential winners.

Tie-Breaking: If there is a tie, tie-breaking criteria will be applied to determine the winning Solutions starting with the Rigor criteria, followed by the Quality of Insights, followed by the Originality and Innovation, then Rigor in such order, as needed. (See Rule 11 "Judging Criteria" below for category descriptions.)

11. **Judging Criteria:** All submissions will be reviewed for content, and no guarantee is made as to when/if any particular Solution will be used or developed by Sponsor. Submissions containing prohibited content and/or otherwise deemed by Sponsor, in its sole discretion, to be inappropriate will be disqualified. Judges' decisions will be final and binding in all matters relating to the Challenge. Challenge Solutions (and all components thereof) will not be returned. By participating, you agree to abide by these Official Rules, the MindSumo Terms of Use and the decisions of the Sponsor and judges, which shall be final and binding in all matters relating to the Challenge.

Category	Max Points
A. **Originality and innovation**	

a) Uniqueness of the proposed method. Does something similar already exist?	10
B. Rigor	
a) Are both the code and results reproducible? (Model sensitivity to multithreading and/or non-deterministic nature of methods notwithstanding.)	20
b) Is the code well commented?	10
c) Is the rationale behind your novel method well thought-out?	20
d) Did you acknowledge and justify all necessary assumptions?	10
e) Is the description of the strengths and limitations of your method well thought-out?	10
C. Quality of insights	
a) Does your method produce both topics and sub-topics?	15
b) Do the topics and sub-topics provide meaningful insights about themes in the corpus? Topic quality will be the average rating of two human judges.	20
c) Is the distribution of topics over documents reasonably stable, when tested in a holdout set?	10
D. Overall quality of presentation	
a) Effective choice of visuals, page layout and use of real estate for both deliverables 2 and 4.	15
b) For the written language parts of the deliverables, does each point flow logically from the previous point?	10
E. Does your method incorporate a sentiment model?	10
F. A quantitative comparison of baseline LDA model to novel method.	10

12. **Prizes:** A total of up to five (5) $3,000 prizes are available. Sponsor reserves the right to award fewer than five (5) prizes in the event an insufficient number of Solutions are received that are of reasonable quality to merit a prize.

13. **Winner Notification and Prize Award:** The potential winners will be notified by Wells Fargo by telephone and/or email beginning about 12/17/19 using the contact information provided MindSumo upon registration. No liability is assumed for any winner notification that is lost, intercepted or not received by a potential winner for any reason. If a potential winner does not reply to the notification within five (5) business days of the date of the notification, s/he may be deemed to be ineligible, at the Sponsor's sole discretion. After responding to the winner notification, potential winners will be sent (based on the domestic U.S. address they provide Wells Fargo) from the Administrator, The Barrett Group, Inc. ("Administrator"), by overnight mail an Affidavit of Eligibility, Liability Release, Assignment of Intellectual Property Ownership, Publicity Release (where legal), Non-Disclosure Agreement, and an IRS W-9 Tax Form ("Prize Claim Documents"). Potential prize winners will be required to sign and notarize the Affidavit. If the Prize

Claim Documents are not properly executed and returned within ten (10) business days of the date on the accompanying letter, the Prize may be forfeited and may be awarded to the next highest scoring Student. Prize Claim Documents will only be sent to an address in the U.S. If a potential winner is outside of the U.S. at the time of winner notification, an extension to properly execute and return the Prize Claim Documents may be granted at Sponsor's sole discretion. Such an extension may not exceed 45 days from the date of initial winner notification. Prize money will be awarded in the form of a check and will be sent via traceable mail to an address in the U.S. within 10 business days of completion of winner verification. Winners will receive an IRS Form 1099 reflecting the value of the Prize received.

Wells Fargo reserves the right to conduct a background check of any and all records of any or all potential winning Students, including without limitation, civil and criminal court records and police reports. To the extent necessary under law, the potential winning Student agrees to authorize this background check. Wells Fargo may also take all steps necessary to corroborate any information provided by the potential winning Student in his/her background check. In that regard, the potential winning Student will be obligated to provide necessary contacts and information so that Wells Fargo may conduct such investigation. Wells Fargo reserves the right (at its sole discretion) to disqualify any potential winning Student from receiving a Prize based on the results of the background check if, in its sole discretion, Wells Fargo determines that awarding a prize to the Student might reflect negatively on Wells Fargo's image and/or reputation. If a potential winner has engaged in or does engage in any conduct that could damage the reputation or business of Sponsor, as determined by Sponsor in its discretion, the potential winner may be disqualified and the prize may be awarded to an alternate winner.

14. **Publicity/Usage Rights:** By participating in this Challenge, each winner irrevocably grants Sponsor and their agents and persons acting with Sponsor's authority, a royalty-free, exclusive, fully transferable, assignable, and sublicensable right and license to use, reproduce, modify, display, transmit, adapt, publish, translate, creative derivative works from, and distribute any submission (or any portion thereof), throughout the world and in perpetuity, in all media now known or hereafter devised and for all lawful business purposes of Sponsor and its affiliates (including without limitation, advertising, promotional, marketing, commercialization, and any other purposes). Each winner agrees to irrevocably assign and transfer to the Sponsor any and all rights, title and interest in the winning Solution including, without limitation, all copyrights and waive all moral rights. Except where prohibited, participation in the Challenge and/or receipt or use of any prize by a winner further constitutes such winner's consent for use of his/her name, likeness, voice, and photo submitted as part of the entry process, for promotional or advertising purposes in any media, worldwide, now known or hereafter developed, in perpetuity, without further review, notification, payment, or consideration (unless prohibited by law). **If you do not agree to the terms of these Official Rules or the MindSumo Terms of Use, do not enter this Challenge.** If any term or condition of these Official Rules conflicts with or is inconsistent with any term or condition of the MindSumo Terms of Use, the terms and conditions in these Official Rules shall govern and be controlling.

15. **Additional Rules:** Entry materials/data that have been tampered with or altered are void. No responsibility is assumed for: (i) lost, late, misdirected, damaged, garbled, incomplete or illegible entries, all of which are void; (ii) error, omission, interruption, deletion, defect, delay in operations or transmission, theft or destruction or unauthorized access to or alterations of entry materials, or

for technical, network, telephone equipment, electronic, mobile device, computer, hardware or software malfunctions of any kind, or inaccurate transmission of or failure to receive entry information by MindSumo or Sponsor on account of technical problems; or (iii) any injury or damage to entrant's or any other person's mobile device or computer related to or resulting from participating in the Challenge. The date and time of entry receipt shall be determined by the official clock on the Challenge Website. Submitted materials become the sole property of Sponsor and will not be acknowledged or returned. If for any reason this Challenge is not capable of being run as planned due to tampering, unauthorized intervention, fraud, technical failures, or any causes beyond the control of the Sponsor which corrupt or affect the administration, security, fairness, integrity or proper conduct of this Challenge, or for any other causes beyond the Sponsor's control, or for any reason Sponsor deems it necessary, Sponsor reserves the right in its sole and unequivocal discretion to cancel, terminate, modify, delay or suspend the Challenge. If the Challenge is cancelled, all submitted Solutions up to the point of termination will be judged by the judges according to the Challenge criteria and winning Solutions will be selected if there are prize-worthy Solutions. By participating in this Challenge, you consent to allow MindSumo to use and to disclose your personal information to Wells Fargo and the Administrator only for the purpose of administering this Challenge.

16. **Release of Liability:** By participating, you release and hold harmless Wells Fargo, MindSumo, and the Administrator and their respective parents, subsidiaries, affiliates, related companies, directors, officers, employees and agents (collectively, the "Releasees") from any and all liability for any injuries, loss or damage of any kind arising from, or in connection with, this Challenge, the prizes or the awards. You agree to indemnify and hold Releasees harmless from any and all claims, damages, loss, injury, expenses and costs (including reasonable attorneys' fees) brought or asserted by any third party against any of the Releasees due to or arising out of your Solution or participation in this Challenge including, but not limited to, trademark, copyright, or other intellectual property rights, right of publicity, or right of privacy or defamation, injury, loss or damage caused or claimed to be caused by your participation in the Challenge and/or the acceptance, awarding, receipt, use and/or misuse of the prize, and you agree that these entities are not responsible for any warranty, representation, or guarantee, expressed or implied, in fact or in law, relating to the Challenge and/or the awarding of any prize except as expressly provided in these Official Rules. By participating, you agree that: (a) any and all disputes, claims, and causes of action arising out of or in connection with the Challenge, or any prizes awarded, shall be resolved individually without resort to any form of class action; (b) any claims, judgments and awards shall be limited to actual out-of-pocket costs incurred, including costs associated with entering the Challenge, but in no event attorney's fees; (c) under no circumstances will any entrant be permitted to obtain any award for, and entrant hereby waives all rights to claim, punitive, incidental or consequential damages and any and all rights to have damages multiplied or otherwise increased and any other damages, other than damages for actual out-of-pocket expenses; and (d) these Official Rules and all disputes shall be governed by the internal laws of the state of California; without regard to principles of conflicts of laws, any disputes shall be resolved exclusively in the state or federal courts in the state of California, and you expressly consent to the jurisdiction of said courts and waive any objection thereto. *For New Jersey residents, the limitations set forth above are inapplicable where attorneys' fees, court costs, or other damages are mandated by statute*.

17. **Privacy Policy:** Any information you submit is subject to Sponsor's privacy policy available at https://www.wellsfargo.com/privacy-security/. Your participation in this Challenge constitutes your consent for Sponsor, MindSumo and Administrator to use your personal information only for the purpose of administering the Sweepstakes and awarding the prizes.

ANY ATTEMPT TO CHEAT OR ANY ATTEMPT TO DELIBERATELY DAMAGE ANY WEBSITE OR UNDERMINE THE LEGITIMATE OPERATION OF THE CHALLENGE IS A VIOLATION OF CRIMINAL AND CIVIL LAWS. SHOULD SUCH AN ATTEMPT BE MADE, THE SPONSOR RESERVES THE RIGHT TO SEEK DAMAGES OR OTHER REMEDIES FROM ANY SUCH PERSON(S) RESPONSIBLE FOR THE ATTEMPT TO THE FULLEST EXTENT PERMITTED BY LAW.

18. **Force Majeure:** The failure of Sponsor, MindSumo and Administrator to comply with any provision of these Official Rules due to an act of God, hurricane, war, fire, riot, earthquake, terrorism, act of public enemies, actions of governmental authorities — outside of their control (excepting compliance with applicable codes and regulations) — or other force majeure event will not be considered a breach of these Official Rules.

19. **List of Winners:** For a list of each winning Student's first name, last initial, city and state, send a self-addressed, stamped envelope to be received by 3/31/20 to: Wells Fargo Challenge Winners, 2554 Lincoln Blvd., PMB 566, Venice, CA 90291-5082.

20. **Sponsor:** Wells Fargo Bank, N.A., 301 South Tryon St., MAC D1130-104, Charlotte NC 28282.

21. **Administrator:** The Barrett Group, Inc., 2554 Lincoln Blvd., PMB 683, Venice, CA 90291-5082

ACCEPTANCE OF OFFICIAL RULES – By clicking "**I agree**" below, you agree to each of the terms and provisions of these Official Rules and the MindSumo Terms of Use and agree on behalf of yourself and your personal representatives, heirs, next of kin, successors and assigns not to contest the validity of this Challenge or any part of it due to the fact you checked the box below. **If you do not agree to all of the terms of these Official Rules or the MindSumo Terms of Use, you understand that you will not be allowed to participate in the Challenge.**

By submitting a Solution in this Challenge, I signify that I agree to these Official Rules and the MindSumo Terms of Use.

    ☐   I agree   ☐ I do **NOT** agree and as such do not wish to be entered into the Challenge