**CIS 662 Introduction to Machine Learning**

*(Mondays & Wednesdays 12:45PM - 2:05PM, Falk 200)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Instructor* | | **Chilukuri K. Mohan** | |  | |
|  | | 4-283 CST, 315 443-2322, [mohan@syr.edu](mailto:mohan@syr.edu) | |  | |
| *Office Hours* | Tuesdays & Thursdays 11AM-12PM <https://syracuseuniversity.zoom.us/j/5663347508?pwd=OFZTa011bno4MTBKQU5MdDhJMDVqdz09>  Meeting ID: 566 334 7508 Passcode: 502152 | | | |
| *Teaching Assistant* | | | **Shivakumar Suresh**  Email: [shsuresh@syr.edu](mailto:shsuresh@syr.edu)  Office hours: Tuesday & Thursday 4PM-5PM, Friday 1PM to 2:30PM  <https://syracuseuniversity.zoom.us/j/3503531690?pwd=K3BPZTZZSE4yVE5UVS9QZnoySmJnQT09>  Meeting ID: 350 353 1690 Passcode: 979761 | |

*Grader:* **Yashpreet Singh Khambay** Email: [ykhambay@syr.edu](mailto:ykhambay@syr.edu)

## Course Description: Linear regression, logistic regression, classification, clustering, and tree-based machine learning; feature extraction and selection; bias-variance trade-off; probabilistic and statistical analyses of learning models and algorithms. Programming assignments.

## Additional Course Description: Data preprocessing; neural networks.

## Prerequisites:

* Familiarity with Differential Calculus, Matrix Algebra, Probability Distributions, Graphs and Trees—needed to understand the foundational principles of ML.
* Ability to program in Python or Julia—needed for HW assignments.
* Ability to write readable and clear sentences in English—needed to answer quizzes and exam questions.

## Learning Objectives:

At the end of the course, students will be able to:

1. Explain the mathematical basis for machine learning (ML)
2. Analyze data and extract features to prepare for the application of ML
3. Distinguish between various classes of problems to which ML may be applied
4. Determine how best to evaluate performance of ML on a given problem
5. Distinguish between various ML models and algorithms
6. Identify which ML models and algorithms are useful for a given problem
7. Apply ML software to solve simple classification and regression problems

## Required Texts / Supplies:

* Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, and Jonathan Taylor, 2023, *An Introduction to Statistical Learning: with Applications in Python*, Springer. (Free electronic copy at <https://www.statlearning.com/>)

### Useful Additional Materials:

* <https://www.datacamp.com/blog/introduction-machine-learning-julia>
* <https://juliapackages.com/c/machine-learning>
* Aurelien Geron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow 3rd Edition, Oct. 2022.
* David Forsyth, 2019, *Applied Machine Learning*, Springer.
* Kevin Patrick Murphy, 2022, *Probabilistic Machine Learning: An Introduction*, MIT Press. (Electronic copy of pre-publication draft version can be downloaded for free from <https://probml.github.io/pml-book/book1.html>)

## Grading:

*Each of the following will be graded out of 4.0 (not using percentages); there is no course project and no mid-term exam.*

* 30% Assignments (9); the last two assignments may be substituted by an [optional] project
* 30% Quizzes (about 4-8 unannounced quizzes; closed-book, open-notes)
* 30% Final exam (Dec. 14, Thu., 12:45-2:45PM; closed-book, open-notes)
* 10% Class participation (open eyes, closed laptop, closed mobile phone)

**Grading Table**

|  |  |
| --- | --- |
| Grades\* | Grade points/credit\* |
| A | 3.67-4.0 |
| A- | 3.34-3.66 |
| B+ | 3.01-3.33 |
| B | 2.67-3.0 |
| B- | 2.34-2.66 |
| C+ | 2.01-2.33 |
| C | 1.67-2.0 |
| C- | 1.34-1.66 |
| F | 0.00-1.33 |

**Course-specific Policies:**

* Course grading is on an **absolute (not relative) scale**—no “curving” based on others’ grades. Your grade will not worsen because someone else does well.
* **Assignments:**
  + Discuss concepts, problems, solutions, and assignments with each other, but **do not look at (or copy) code from other students**.
  + You may reuse (and build on) publicly available and well-tested code from Github or other reliable sources (ChatGPT is not reliable enough), as long as the submission clearly indicates what was taken from where, **giving credit to the original author(s)** or software generating the code; **not doing so will be considered a violation of academic integrity**, resulting in F grade in the course and reporting to the college and university.
  + Assignments must be submitted on Blackboard **by 11:59PM** on the due date. Late submissions will be penalized at the rate of **10% per day.**
  + Use Pluto (see <https://plutojl.org/>) or JuPyTeR notebooks.
  + All assignments carry equal weight, even though some are easier than others.
* **Quizzes and Final Exam:**
  + Quizzes and examinations are **closed-book, open-notes,** with no communication permitted—during a quiz/exam, students are forbidden from speaking and from looking at any laptop, tablet computer, or mobile phone.
  + Students may refer to any notes in their own handwriting—**no printed /photocopied materials** are permitted; these notes must be placed on the desk in front of the student before the quiz/exam begins.
  + Seating for quizzes and exams will be **assigned by the instructor**.
  + **Silence** is expected in quizzes/exams until all students complete their quiz/exam submissions; speaking to other students during a quiz/exam will be considered a violation of academic integrity.
  + Quizzes are **unannounced**. The material tested in a quiz would be covered in the preceding two weeks. Utilize the weekend to ensure that you understand the material covered during the week.
  + There will be **no make-up quizzes**. Email [mohan@syr.edu](mailto:mohan@syr.edu) at least an hour BEFORE the class begins if you will be absent from class for a legitimate reason e.g., travel, illness, or a job interview (but not due to waking up late) your average grade from other quizzes will be used to fill in the missing data when the course grade is assigned.
  + **Turnitin** now includes a built-in AI detection tool, and may be used in this course along with other tools to help detect academic integrity violations.
* **Class participation**:
  + To receive 4.0 for this (10% of the course grade), you **must pay attention** in all classes, **ask and answer questions**, and avoid distractions (laptop, mobile phone, tablet).
  + Your grade for this will be 0.0 **if you miss multiple classes** for any reason, do **not pay attention** during classes (e.g., if you open up your laptop, use your mobile phone, work on other courses, chat with friends, don’t ask/answer questions, etc.). Note that if you receive 4.0 in every quiz, assignment, and final exam., but score 0.0 for class participation, your overall course grade would be 3.6 (i.e., A-).
  + I strongly recommend keeping a **physical notebook** in which you take hand-written notes during classes, which you are free to consult during quizzes and the final exam.

***Tentative Class Schedule:***

| Date | Topic | Details |
| --- | --- | --- |
| 8/28, 8/30, 9/6 | Introduction | What is ML; Learning in humans; Difference between ML algorithms and models; Local search vs. global search vs. ensembles vs. interacting experts |
| 9/11 | Data preprocessing | Normalization, Winsorizing, anomalous values, missing data, nominal/discrete attributes, encoding, embedding |
| 9/13 9/18 | Feature selection and input dimensionality compression | Features as combinations of attributes; Assessing the usefulness of a feature; Principal Component Analysis (PCA); Nonlinear attribute combinations; Iterative leave-one-out feature selection |
| 9/20, 9/25 | Evaluation measures | ML as optimization; Mean squared error, accuracy, confusion matrix, F1, and other cost/performance measures; imbalanced sets, asymmetric cost functions, incommensurate dimensions |
| 9/25, 9/27 | Validation | Training vs. test data; cross-validation; over-training; generalization; over-parameterization; bias-variance tradeoff; regularization |
| 10/2 | Models | Equations, Trees, Networks |
| 10/4 | Model-free approaches | Interpolation, k-Nearest Neighbors, Case-based reasoning |
|  | Fall break |  |
| 10/11 | Learning Algorithms | Different kinds of ML algorithms: Unsupervised, supervised, semi-supervised, reinforcement, … |
| 10/16 | Unsupervised learning | Clustering; specialization; competitive learning; association rules; data mining |
| 10/18 | Shallow Neural Networks | Perceptrons, Feedforward networks with 1-2 hidden layers |
| 10/23, 10/25 | Regression | Linear and Logistic Regression; Support Vectors |
| 10/25, 10/27 | Deep Learning | Deep network models; Encoder-Decoder model; Attention; Transformers; Large Language Models |
| 11/1 | Decision trees | CART, C4.5, oblique trees |
| 11/3 | Ensembles | Random forests, Mixture of Experts, Decision fusion |
| 11/8, 11/10 | Boosting and Bagging | Adaboost, Gradient boosting, LASSO |
| 11/15, 11/17 | Probabilistic ML | Naïve Bayesian; ML for causal networks and uncertainty propagation |
|  | Thanksgiving | Break |
| 11/27, 11/29 | Reinforcement learning (RL) | Discrete vs. Continuous RL; Q-Learning; Neural & Deep RL |
| 12/4 | Catchup/spillover | Above topics, as needed |
| 12/6 | Societal issues | Ethics, Fairness, Bias in ML; the “Singularity”; etc. |
| 12/11 | Conclusion | Review |

**Homework Assignment Schedule**

**HW0 (due 9/13) [Data Extraction]**

1. Recall your course-specific PIN ( a two-digit number). Select the corresponding university  (with rank same as the last two digits of your PIN) from the list at <https://www.topuniversities.com/where-to-study/north-america/united-states/ranked-top-100-us-universities>; if there's a tie, treat them in the order given, e.g., two universities are ranked #30, so use #30 for the first and #31 for the second.
2. Look up the webpage of the selected university's computer science department (if there's no CS department, use the most similar department, e.g., EECS or CSE or CE).
3. Find ten of the current faculty members whose ranks are Assistant Prof., Associate Prof., Prof., or Distinguished Prof. (Don't select those whose titles are Lecturer, Adjunct Prof., Emeritus Prof., Teaching Prof., or Prof. of Practice.), whose first & last name initials are not identical.
4. Go to the Google Scholar Citations page for each of those ten faculty members, and list the following information for each: univ. rank, first initial, last initial, numbers of citations in each of the years 2017-2022, h-index, and i-10 index.  For example, if your PIN is 22, then you may have the following entry for Prof. Lav Varshney from Univ. of Illinois: 22, L, V, 633, 688, 732, 861, 946, 1054, 35, 100
5. Combine all ten records into a CSV file with the name of the university, e.g., Illinois.csv
6. Generate a noisy version of this file, with a different filename preceded by "noisy", e.g., noisyIllinois.csv, in which each of the citation counts is randomly modified by adding an integer in the range [-10, +10],  e.g., the previous entry may be modified to 22, L, V, 631, 678, 742, 866, 942, 1050, 35, 100
7. Generate a missing data version of the noisy file, with a different filename preceded by "missing", e.g., missingnoisyIllinois.csv, randomly omitting one of the citation numbers in each record, e.g., the previous entry may be modified (omitting the third citation number entry) to 22, L, V, 631, 678, , 866, 942, 1050, 35, 100
8. Submit all three files.

**HW1 (due 9/27) Missing Values (Tentative):**

1. Use MICE (Multiple Imputation by Chained Equation) or another sophisticated approach for missing data imputation, e.g., see  <https://www.geeksforgeeks.org/missing-data-imputation-with-fancyimpute/> , to impute the missing values in the dataset you obtained and modified.  Evaluate the results in terms of the average error (i.e., difference between the actual value and the imputed value) over all the ten data points (faculty members).
2. Try the following simpler alternatives:

(a) Individual\_mean: Replace by the mean over all years for which data is available for that faculty member.

(b) Individual\_median: Replace by the median over all years for which data is available for that faculty member.

(c) Field\_mean: Replace by the mean over all faculty members for which data is available for that year.

(d) Field\_median:  Replace by the mean over all faculty members for which data is available for that year.

(e) Local\_gradient: For missing data in the middle years (2018-2021) Replace by the average of the preceding and following years, for that faculty member.   For missing data in 2017, use twice the value of 2018 minus the value of 2019.   For missing data in 2022, use twice the value of 2021 minus the value of 2020.

(f) Nearest\_neighbor\_L1: For each of the other faculty members, calculate an averaged  distance from each of the others, based on fields for which both faculty members have values, e.g., the distance between (..., n17, , n19, ,n20, n21, n22, ...) and (..., m17, m18, ,m20, m21, m22, ...) = (|n17-m17| + |n20-m20|  + |n21-m21| + |n22-m22|)/4.  For each faculty member, compute the nearest other faculty member using this distance measure, for whom the missing value is not in the same field, and replace the missing value by that field value for the nearest neighbor.

(g) Nearest\_neighbor\_L2: Similar to the above but use the Euclidean distance measure, e.g., the distance between (..., n17, , n19, ,n20, n21, n22, ...) and (..., m17, m18, ,m20, m21, m22, ...) = ((n17-m17)^2 + (n20-m20)^2  + (n21-m21)^2 + (n22-m22)^)/4.

3. Which approach has worked best: MICE, or any of the above simpler alternatives?

4. What is the possible explanation?

**HW2 (due 10/11) Clustering**

Use an appropriate distance measure, to determine nearest neighbors, and to group individuals in the data set based on all the 2017-2021 citation columns in the data set.

Compare the following two predictions for the 2022 citation numbers:

(1) same as the 2022 citation number of the nearest neighbor (based on prior year citations);

(2) average of the three nearest neighbors;

(3) average of all others in the same cluster.

**HW3 (due 10/18) Shallow Neural Networks**

Try a 1-hidden layer neural network (5-3-1 architecture).

Evaluate, and compare with HW2.

**HW4 (due 11/1) Regression**

**Linear regression:**  Fit a line to go very near the 2017-2021 citation columns, minimizing MSE.  Use that line to predict the 2022 citation numbers.

**Logistic regression:** Similar to the above, but use a logistic function.

How do the results compare with those of HW2 and HW3?

**HW5 (due 11/8) Support Vectors**

Try Support Vector Regression; compare with the previous.

**HW6 (due 11/15) Classification with a neural network**

Instead of estimating a value, let us cast the data into a classification context.

Place individuals into 3 categories.

1. No/very little increase (<5% per year) in citation numbers over time.
2. Steady increase (6%-15% per year)
3. Substantial increase (>15% per year)

Generate a new column for the data set, based on the above class labels.

Use a 1-hidden layer 6-6-3 neural network to solve this classification problem, and evaluate the results.

**HW7 (due 11/29) Random forests**

Try a random forest approach for classification.

(It probably won't perform well; say why.)

Introduce 5 new features based on the citation numbers, and use them in the RF instead of the citation numbers directly.  Each new feature is

((citation number in year n+1)-(citation number in year n))/(citation number in year n)

for 2016<n<2022.

**HW8 (due 12/6) Bias Variance Tradeoff Analysis**

Illustrate how the bias-variance tradeoff plays in the results you obtained, varying the relative percentage of data used in training.  Show the results in a graph, with the y-axis showing the MSE or accuracy on the test data, and the x-axis showing the percentage of data used as test data, in the range 5% to 50%.

**Optional Project (due 12/15): Students may choose to work on a project instead of HW7 and HW8. If you decide to do so, please submit a 1-page project proposal by Nov. 17, clearly describing what you plan to do.**

**The material below is extracted from** <https://academicaffairs.syracuse.edu/important-syllabus-reminders/> **.**

**ACADEMIC DROP DEADLINE**

As part of our efforts to track satisfactory academic progress, the Academic Drop Deadline and the Financial Drop deadline will both occur on September 18, 2023, for the fall semester and February 6, 2024, for the spring semester. Students may still withdraw from courses after these deadlines; this would place a ‘WD’ grade on their transcripts. Students enrolled in “flex” classes (Flexibly formatted classes) have different deadlines and will need to check MySlice for the academic drop deadline that pertains to their class.

**Academic Integrity**

Syracuse University’s [Academic Integrity Policy](https://class.syr.edu/academic-integrity/policy/) reflects the high value that we, as a university community, place on honesty in academic work. The policy holds students accountable for the integrity of all work they submit and for upholding course-specific, as well as university-wide, academic integrity expectations. The policy governs citation and use of sources, the integrity of work submitted in exams and assignments, and truthfulness in all academic matters, including course attendance and participation. The policy states that any work a student submits for a course must be solely their own unless the instructor explicitly allows collaboration or editing. The policy also requires students to acknowledge their use of other peoples’ language, images or other original creative or scholarly work through appropriate citation. These expectations extend to the new, fast-growing realm of artificial intelligence (AI) as well as to the use of websites that charge fees or require uploading of course materials to obtain exam solutions or assignments. Students are required to ask their instructor whether use of these tools is permitted – and if so, to what extent – before using them to complete any assignment or exam. Students are also required to seek advance permission from instructors if they wish to submit the same work in more than one course. Failure to receive this permission in advance may violate the Academic Integrity Policy.

Upholding Academic Integrity includes abiding by instructors’ individual course expectations, which may include the protection of their intellectual property. Students should not upload, distribute, or otherwise share instructors’ course materials without permission. Students found in violation of the policy are subject to grade sanctions determined by the course instructor and non-grade sanctions determined by the School or College where the course is offered, as outlined in the Violation and Sanction Classification Rubric. Students are required to read an online summary of the University’s academic integrity expectations and provide an electronic signature agreeing to abide by them twice a year during pre-term check-in on MySlice.

**Course Failure as a Grade Sanction for Level 1 Violations**

The Violation and Sanction Classification Rubric establishes recommended guidelines for the determination of grade penalties by faculty and instructors, while also giving them discretion to select the grade penalty they believe most suitable, including course failure, regardless of violation level. Any established violation in this course may result in course failure regardless of violation level.

Using websites that charge fees or require uploading of course material (e.g., Chegg, Course Hero) to obtain exam solutions or assignments completed by others and present the work as your own violates academic integrity expectations in this course and may be classified as a Level 3 violation, resulting in suspension or expulsion from Syracuse University.

This class will use the plagiarism detection and prevention system Turnitin. You will have the option to submit your papers to Turnitin to check that all sources you use have been properly acknowledged and cited before you submit the paper to me. I may also submit papers you write for this class to Turnitin, which compares submitted documents against documents on the Internet and against student papers submitted to Turnitin at Syracuse University and at other colleges and universities. I will take your knowledge of the subject matter of this course and your writing level and style into account in interpreting the originality report. Keep in mind that all papers you submit for this class will become part of the [Turnitin.com](http://turnitin.com/) reference database solely for the purpose of detecting plagiarism of such papers.”

**ATTENDANCE POLICY (including Absence Notification)**

Attendance in classes is expected in all courses at Syracuse University. For more information visit [Information for Students: Non-attendance or Stopped Attending](http://registrar.syr.edu/students/non-attendance/)

If a student is unable to participate in-person or virtually for an extended period of time (48 hours or more), the student may request an absence notification from their home school/college Dean’s Office or through Student Outreach and Retention (SOaR) office. Instructors will be notified via the “Absence Notification” flag in Orange SUccess.

Barnes Center at the Arch (Health, Counseling, etc.) staff will not provide medical excuse notes for students. When Barnes Center staff determine it is medically necessary to remove a student from classes, they will coordinate with SOaR case management staff to provide appropriate notification to faculty through Orange Success. For absences lasting less than 48 hours, students are encouraged to discuss academic arrangements directly with their faculty.

Additional information may be found at [Student Outreach and Retention: Absence Notifications](https://experience.syracuse.edu/soar/support-services/absence-notifications/)

**BLACKBOARD LEARNING MANAGEMENT SYSTEM**

This class will use the Blackboard Learning Management to house the syllabus, course content, links to external course materials, assignments, quizzes, exams, feedback, and grades. Note when submitting materials that the University’s Blackboard Learning Management System is on Eastern Time.

Information about Blackboard is available on [Answers Blackboard](https://answers.syr.edu/display/blackboard01/Blackboard#space-menu-link-content); alternatively, you can contact Information Technology Services by sending an email to [help@syr.edu,](mailto:help@syr.edu) calling 315.443.2677, or in-person at the ITS Service Center, located at 1-227 CST in the Life Sciences Complex. Business hours for the Service Center can be found on the ITS Website at <http://its.syr.edu/its_service_center/>

**Accessible Blackboard Content**

Syracuse University is using Blackboard Ally, a tool to help enhance the usability and accessibility of course documents in the Blackboard learning management system. Ally provides students with multiple accessible formats of the original document to select those that fit their unique needs. Currently, Ally offers accessible versions of Portable Document Format (PDF) files, Microsoft Office files (Word and PowerPoint), images and uploaded HTML files. Students will see a clickable “A,” which pops up in a window, and they can select one or more accessible alternative documents to download and use.

Additional information on the Blackboard Ally tool is available on [Answers Blackboard Ally](https://answers.syr.edu/display/blackboard01/Blackboard%2BAlly); alternatively, you can contact Information Technology Services by sending email to [help@syr.edu](mailto:help@syr.edu) or calling 315.443.2677.

**DISABILITY SYLLABUS STATEMENT**

Syracuse University values diversity and inclusion; we are committed to a climate of mutual respect and full participation. There may be aspects of the instruction or design of this course that result in barriers to your inclusion and full participation in this course. I invite any student to contact me to discuss strategies and/or accommodations (academic adjustments) that may be essential to your success and to collaborate with the Center for Disability Resources (CDR) in this process.

If you would like to discuss disability-accommodations or register with CDR, please visit [Center for Disability Resources](https://disabilityservices.syr.edu/). Please call (315) 443-4498 or email [disabilityresources@syr.edu](mailto:disabilityresources@syr.edu) for more detailed information.

The CDR is responsible for coordinating disability-related academic accommodations and will work with the student to develop an access plan. Since academic accommodations may require early planning and generally are not provided retroactively, please contact CDR as soon as possible to begin this process.

**DISCRIMINATION OR HARASSMENT**

Federal and state law, and University policy prohibit discrimination and harassment based on sex or gender (including sexual harassment, sexual assault, domestic/dating violence, stalking, sexual exploitation, and retaliation). If a student has been harassed or assaulted, they can obtain confidential counseling support, 24-hours a day, 7 days a week, from the [Sexual and Relationship Violence Response Team](https://ese.syr.edu/bewell/sexual-and-relationship-violence/) at the Counseling Center (315-443-8000, Barnes Center at The Arch, 150 Sims Drive, Syracuse, New York 13244). Incidents of sexual violence or harassment can be reported non-confidentially to the University’s Title IX Officer (Sheila Johnson Willis, 315-443-0211, [titleix@syr.edu](mailto:titleix@syr.edu), 005 Steele Hall). Reports to law enforcement can be made to the University’s Department of Public Safety (315-443-2224, 005 Sims Hall), the Syracuse Police Department (511 South State Street, Syracuse, New York, 911 in case of emergency or 315-435-3016 to speak with the Abused Persons Unit), or the State Police (844-845-7269). I will seek to keep information you share with me private to the greatest extent possible, but as a professor I have mandatory reporting responsibilities to share information regarding sexual misconduct, harassment, and crimes I learn about with the University’s Title IX Officer to help make our campus a safer place for all.

**FAITH TRADITION OBSERVANCES**

[Syracuse University’s Religious Observances Policy](https://policies.syr.edu/policies/university-governance-ethics-integrity-and-legal-compliance/religious-observances-policy/) recognizes the diversity of faiths represented in the campus community and protects the rights of students, faculty, and staff to observe religious holy days according to their traditions. Under the policy, students are given an opportunity to make up any examination, study, or work requirements that may be missed due to a religious observance, provided they notify their instructors no later than the academic drop deadline. For observances occurring before the drop deadline, notification is required at least two academic days in advance. Students may enter their observances in MySlice under Student Services/Enrollment/My Religious Observances/Add a Notification.

**HEALTH & WELLNESS CONSIDERATIONS**

Mental health and overall well-being are significant predictors of academic success. As such it is essential that during your college experience you develop the skills and resources effectively to navigate stress, anxiety, depression, and other mental health concerns. Please familiarize yourself with the range of resources the Barnes Center provides (<https://ese.syr.edu/bewell/>) and seek out support for mental health concerns as needed. Counseling services are available 24/7, 365 days, at 315-443-8000, and I encourage you to explore the resources available through the Wellness Leadership Institute, <https://ese.syr.edu/bewell/wellness-leadership-institute/>

With COVID-19 still circulating, I recommend everyone to wear a mask when they are in close proximity to others, e.g., in a classroom. When meeting me in my office, you must wear a mask.

**POLICY ON FACULTY USE OF STUDENT ACADEMIC WORK**

Educational use of student work: I may use academic work that you complete this semester for educational purposes in this course during this semester. Your registration and continued enrollment constitute your permission.

Educational use of student work: I may use academic work that you complete this semester in subsequent semesters for educational purposes. Before using your work for that purpose, I will render the work anonymous by removing identifying material.