Tinashe Michael Tapera, MSc

1022 S. 47th Street ● Philadelphia, PA 19143 ● 267-441-7206
Email: tinashemtapera@gmail.com
Website: tinashemtapera.com

SUMMARY

A recent psychology graduate-turned data scientist — my mission is to understand and predict human behaviour using advanced data mining and machine learning tools. Expert R and proficient python programmer, with experience applying machine learning in the domains of clinical psychology, clinical neuroscience, and human resource (HR) psychology.

EDUCATION

Drexel University — September 2016 - June 2018
Accelerated Master of Science in Psychology — Philadelphia, PA
Behavioral Data Analysis Concentration
Graduate Minor in Computer Science
Graduating GPA: 3.54 (Cum Laude)

Drexel University — September 2013 - June 2017
Bachelor of Science in Psychology — Philadelphia, PA
Graduating GPA: 3.51 (Cum Laude)

St. John's College — 2007 - 2012
International Baccalaureate Diploma Program — Harare, Zimbabwe
Higher level: Physics, Business Studies, English
Standard level: Mathematics, Chemistry, French
32 points

HONORS AND AWARDS

Drexel University

Statistical and Mathematical Sciences Institute Undergraduate Travel Award	2016
College of Arts & Sciences Humanities Fellowship in Psychology	2015
Dean's List	Multiple
A. J. Drexel Academic Scholarship	2013 - 2018

St. John's College

Athene Cup for Cultural Person of the Year	2012

PROFESSIONAL ORGANIZATIONS

American Statistical Association (ASA) Member and Mentee
Eastern North American Region of the International Biometrics Society (ENAR) Member
Society of Behavioral Medicine (SBM) Member
Phi Eta Sigma National Honors Society Member
Drexel University Gospel Choir Member

WORK EXPERIENCE

Brain & Behavior Laboratory October 2018 - Present
Neuroimaging Data Analyst Philadelphia, PA
Supervisor: Theodore Satterthwaite, PhD

- Use programming tools and technologies to curate, analyse, and present brain scan imaging data

- Developed data pipelines to transfer multiple data sets (10+ terabytes) between data warehouses in Python and Bash.
- Contributed R code/analysis and Rmarkdown documentation to development of a novel approach for modelling spatially varying associations between multi-modal images.
- Preprocess and analyse neuroimaging data using common software tools (fMRIprep, XCPengine, multi-voxel pattern analysis).
- Developed and actively maintain two data curation software packages in Python:
  - bids-on-flywheel
  - fw-heudiconv

Salesforce June 2017 - September 2017
Data Science Intern San Francisco, CA
Supervisor: Ernest Ng, PhD

- Focused on discovery of organizational insight using internal human resources data sets.
- Developed a semi-supervised learning algorithm for tracking employee performance by matching topic models in ongoing feedback data and goal-setting data.
- Designed data processing pipelines for analyses in Python, Bash, and MySQL.
- Examined employee feedback patterns using various text mining techniques:
  - Parts-of-speech tagging,
  - Topic modelling,
  - Sentiment analysis.
- Investigated comorbidity of employees' insurance claims data to dynamically classify claim types and profile employees.
- Carried out ad-hoc statistical investigations as requested.

Arzoo LLC October 2015 - April 2016
Private Equity Analyst Philadelphia, PA
Supervisors: Ahmed Makani, Megan Strouss-Rooney, MSW

- Scraped various online business databases to create large datasets of businesses (5000+ cases).
- Developed a smooth data munging pipeline in Excel using advanced functions to cut down data munging time for interns by almost 75%.
- Taught data preparation techniques to other interns using Excel templates and basic shell scripting.
- Analyzed cases using financial criteria and seller profile criteria to identify business owners with high potential to sell.
- Initiated negotiations with business owners by email, mail and by phone.

RESEARCH EXPERIENCE

Quantitative Psychology & Statistics Laboratory 2015 - 2019
Graduate Research Assistant Drexel University
Supervisor: Fengqing Zoe Zhang, PhD

- Investigate advanced data mining and machine learning models for data-intensive research in psychology, neuropsychology, and epidemiology.

- "Momentary Changes in Heart Rate Variability Can Detect Risk for Emotional Eating Episodes."
  - Aimed to predict emotional eating episodes using heart rate data sourced from smartwatches in an emotional eating study.
  - Methods include continuous Heart Rate Variability analysis in the time and frequency domain, feature engineering and selection, and binary classification using support vector machines
  - Achieved moderately high accuracy (77%), sensitivity (79%) and specificity (75%) when predicting the likelihood of an emotional eating episode occurring within the following 30 minute window, based on fluctuations of heart rate.
  - Models can be used to continuously monitor patients at risk of emotional eating episodes, and intervene on positive signals. Manuscript in prep.

- "Application of Advanced Data Mining Models to Identify Dietary Patterns Associated with Risk of Cardiovascular Disease."
  - Data sourced from National Health and Nutrition Examination Survey.
  - Aimed to predict levels of cardiovascular disease biomarkers (e.g. cholesterol) using high-dimensional food and behaviour survey responses.
  - Employed and compared principal components analysis + linear regression, and LASSO regression.
  - Achieved adjusted-R^2 of over 0.85 in predicting individual biomarker levels using LASSO approach.

- "Improved Modelling of Smartphone-based Ecological Momentary Assessment Data for Dietary Lapse Prediction."
  - Ecological Momentary Assessment data sourced via periodic mobile app survey.
  - Aimed to predict when participants in dietary adherence study would lapse (i.e. cheat) on their diet by classifying survey responses into lapse or non-lapse probabilities.
  - Developed a novel algorithm for classification in the case of highly imbalanced classes:
    - Hypothesized a neighborhood-based approach to create small subsets of training data based on similarity to the test set observation.
    - Implemented individual classifiers for each subset and test observation employed ensemble rule evaluation for the resulting predictions.
    - Results show moderate improvements in specificity and sensitivity over existing data-balancing and ensemble rule techniques.

- "Identifying Autism Diagnostic Interview: Revised Algorithm Items that Significantly Distinguish Autism Spectrum Disorder and Down Syndrome."
  - Data gathered through Autism Diagnostic Review questionnaire (ADI-R), including participants diagnosed with Autism Spectrum Disorder, Down Syndrome, or both.

- Aimed to infer critical diagnostic criteria for early detection and differentiation of diagnoses.
- Built both dichotomous and multiclass classification models for prediction of diagnosis, employing methods such as logistic regression, LASSO, and classification & regression trees such as Random Forest.

- "Modelling Zero-Inflated Distributions in Bouted Physical Activity Data."
  - Data sourced from individual fitness trackers to measure amounts of moderate-to-vigorous physical activity.
  - Assumed zero-inflated compound Tweedie Poisson distribution would best model the change in trajectory of participants' moderate-to-vigorous physical activity (MVPA) bouts over time.
  - Achieved MSE of 11.98 using a fixed effects growth curve model to predict MVPA bouts.

SAMSI Interdisciplinary Undergraduate Workshop May 2016
Supervisors: Benjamin Risk, BS and Sujit Ghosh PhD NC State University

- Awarded $900 travel award to participate in the Statistical and Mathematical Sciences Institute Interdisciplinary Undergraduate Workshop 2016.
- Worked with a team of multidisciplinary students to predict multiple sclerosis diagnosis.
- Investigated statistical modelling of Diffusion Tensor Imaging data for MS patients in R and MATLAB.
- Loaded, cleaned, and explored datasets using statistical summaries and visualizations.
- Identified and engineered predictive features for multiple sclerosis diagnosis.
- Successfully modelled MS diagnosis using 3 different logistic regression approaches with 81% classification accuracy.

Laboratory for Innovations in Health-Related Behavior Change 2015
Supervisors: Evan Forman, PhD and Stephanie Goldstein BS Drexel University

- Data Manager: A Companion Smartphone App to Enhance Dietary Adherence through Predictive Machine Learning.
- Data collection, entry, and cleaning using Excel functions (text manipulation, LOOKUPs, case matching).
- Managed participant research and clinical files.
- Conducted in-person and phone assessments for data collection and in-person diagnostic interviewing.
- Independently investigated hypotheses on self-report and baseline measurement discrepancies using SPSS.

PUBLICATIONS

Juarascio, A. S., **Tapera, T. M.**, Crochiere, R., Palermo, M., Zhang, F. Z. (manuscript in preparation). Momentary Changes in Heart Rate Variability Can Detect Risk for Emotional Eating Episodes.

Godfrey, M., **Tapera, T. M.**, Hepburn, S., Fidler, D. J., Zhang, F. Z., Rosenberg, C. R., Lee, N. (manuscript). Autism Spectrum Disorder (ASD) Symptom Profiles of Children with Comorbid Down Syndrome (DS) and ASD: A Comparison with Children with DS-only and ASD-only.

Zhang, F. Z., **Tapera, T. M.**, Goldstein. S.P., & Forman, E.M., (manuscript). Improved Modeling of Smartphone-based Ecological Momentary Assessment Data for Dietary Lapse Prediction.

Zhang, F. Z., **Tapera, T. M.**, & Gou, J.T.,(manuscript). Application of a New Dietary Pattern Analysis Model in Nutritional Epidemiology.

CONFERENCE PRESENTATIONS

Valcarel, A. M., **Tapera, T. M.**, Vanderkar, S. N., Adebimpe, A., Roalf, D., Raznahan, A., Satterthwaite, T., Shinohara, R. T., & Linn, K. A. (2019, May). Approaches for Modelling Spatially Varying Associations Between Multimodal Images. Poster presented at Shape Modeling International Conference 2019, Irvine, CA.

Zhang, F. Z., **Tapera, T. M.**, Goldstein. S.P., Forman, E.M. (2018, March). Improved Modeling of Smartphone-based Ecological Momentary Assessment Data for Dietary Lapse Prediction. Talk presented at the ENAR Spring Meeting 2018, Atlanta, GA.

Tapera, T. M., Zhang, F. Z., Forman, E. (2018, February). Novel Neighbourhood-based Classification Algorithms for Dietary Lapse Detection. Talk presented at WELL Center Symposium 2018, Philadelphia, PA.

Godfrey, M., **Tapera, T. M.**, Zhang, F. Z., Lee, N (2018, February). Identifying Autism Diagnostic Interview- Revised Algorithm Items that Significantly Distinguish Autism Spectrum Disorder and Down Syndrome. Talk presented at the International Neuropsychological Society Conference, Washington, D.C.

Tapera, T. M., Ng, E., (2017, September). Using Natural Language Processing to Link Feedback & V2MOM Data. Talk presented at the Tech & Product Intern Demo Day at Salesforce HQ, San Francisco, CA.

Tapera, T. M., Zhang, F. Z., (2017, March). Application of Advanced Data Mining Models to Identify Dietary Patterns Associated with Risk of Disease. Poster presented at the ENAR Spring Meeting 2017, Washington, D.C.

Tapera, T. M., Goldstein, S.P., Evans, B., & Forman, E.M., (2016, April). Does Ecological Momentary Assessment Data Reflect Baseline Self-Report in Weight Loss Treatment? Poster presented at the 37th Annual Meeting and Scientific Sessions of the Society of Behavioral Medicine, Washington, D.C.

Forman, E.M., Goldstein. S.P., Evans, B., Manasse, S.M., Juarascio, A.S., Butryn, M.L., & **Tapera, T. M.**, (2016, April). A Preliminary Investigation of a Personalized Risk Alert System for Weight Control Lapses. In S.P. Goldstein (Chair), *Harnessing the Power of Predictive Learning to Promote Health Behavior Change: Developing and Testing Novel Technology.* Symposium presented at the 37th Annual Meeting and Scientific Sessions of the Society of Behavioral Medicine, Washington, D.C.

Goldstein, S.P., Evans, B., **Tapera, T. M.**, Forman, E.M., & Butryn, M.L., (2016, April). Is Prompting Problematic?: Considerations for Long-term Ecological Momentary Assessment? Poster presented at the 37th Annual Meeting and Scientific Sessions of the Society of Behavioral Medicine, Washington, D.C.

TEACHING EXPERIENCE

Graduate-level Psychology Statistics Sequence September 2017 - June 2018
Teaching Assistant Drexel University
Supervisor: Karol Osipowicz, PhD

- Tutor 1st year Master's students in fundamental statistical procedures relevant to behavioural and social science research:
  - Central tendency and variability,
  - Probability and expectation,
  - Hypothesis testing,
  - Correlation and regression,
  - Chi-square, ANOVA, and variants,
  - Non-parametric tests.
- Direct laboratory sessions to allow students to practice and develop their skills.
- Host office hours to allow students to address lingering questions and concerns.

VOLUNTEER EXPERIENCE

Peer Counseling Helpline 2014 - 2019
Counselor, Vice President of Scheduling & Communications Drexel University
Supervisor: Scott Sokoloski, PhD

- Chair and record minutes for staff and board meetings.
- Manage ~30 counselors per term and co-ordinate staff shifts and meetings for the term.
- Serve on the board as a counselor liaison.
- Answer calls and counsel students through psychological and emotional distress.
- Analyzed Gmail archive in Python to establish insightful statistics and key performance indicators about the organization.
  - Applied topic modeling algorithms on call logs to classify call topics and caller profiles.

RELEVANT SKILLS AND TOOLS

- Proficient R programming.
- Proficient Python programming.
- Proficient SPSS.
- Novice Bash/shell, MATLAB, AppleScript, MySQL, Java scripting.
- Qualtrics Survey.
- IRB Certified and Compliant.
- Proficient Microsoft Office Suite.

RELEVANT COURSEWORK

Undergraduate

- Biological Basis of Behavior
- Physiological Psychology
- Neuropsychology
- Industrial/Organizational Psychology
- Experimental Psychology
- Psychological Testing & Assessment
- Dietary Behavior & Technology
- Machine Learning in Psychology
- Introduction to Data Science
- Multivariate Data Analysis
- Introduction to Computer Science
- Programming I & II
- Advanced Programming Techniques & Data Structures
- Mathematical Foundations of Computing

Graduate

- Research Methods I & II
- Cognitive Psychology
- Citizen Science
- Behaviour Analysis
- Statistical Data Analysis I&II
- Statistical Data Analysis III: Advanced Topics
- Behavioural Data Mining
- Hierarchical modelling
- Theoretical Computer Science Foundations
- Programming Foundations
- Advanced Programming Techniques

RELEVANT PROJECTS

- fw-heudiconv – Heudiconv-Style BIDS curation of imaging data on Flywheel (Python)
- bids-on-flywheel – a toolkit for querying, editing, and uploading BIDS data to Flywheel (Python)
- Encourage.me – A shiny app that generates random messages of encouragement (R)
- PC Dashboard – An interactive analytics dashboard for real-time analysis and visualisation of Peer Counselling call log data (R, Shiny)
- Quantitative Analysis of the Peer Counseling Helpline Gmail archive using Natural Language Processing techniques (Python);
- Investigating my sleeping habits using the Sleep Cycle app and my sleep data (R);
- Equivalence & Non-Inferiority Testing: an undervalued hypothesis testing approach (R);
- Kaggle Housing Prices Challenge (R);
- Solving the Knight's Tour problem (Python);
- Automating scheduled emails (AppleScript);
- Confirmatory analysis of air quality data in the Philadelphia area (R);
- A semi-supervised learning algorithm for matching text documents from different corpuses (Salesforce internship).

Documentation available at github.com/TinasheMTapera.

REFERENCES

Fengqing Zoe Zhang, PhD
Assistant Professor
MS Research Mentor
Drexel University
3141 Chestnut Street, Stratton 326
Philadelphia, PA 19104
Phone: 215.553.7172
Email: fengqing.zhang@drexel.edu

Nancy Lee, PhD
Assistant Professor
Director of MS and BS/MS Programs
Drexel University
3141 Chestnut Street, Stratton 123E
Philadelphia, PA 19104
Phone: 215.895.2937
Email: nancy.raitano.lee@drexel.edu

Karol Osipowicz, PhD
Assistant Teaching Professor
Graduate Statistics Sequence Professor
Drexel University
3141 Chestnut Street, Stratton 214
Philadelphia, PA 19104
Tel: 215.571.4406
Email: karol.zdislaw.osipowicz@drexel.edu

Scott Sokoloski, PhD
Staff Psychologist
Director, Peer Counseling Helpline
Drexel University
3141 Chestnut Street, Creese 201
Philadelphia, PA 19104
Tel: 215.895.1415
Email: sas464@drexel.edu